perspectiva filosófica

vol. 47, n. 2, 2020

número especial

Filosofia da Lógica

Revista do Departamento de Filosofia
Universidade Federal de Pernambuco

ISSN 2357-9986

EDITORIAL

Em nome da revista "Perspectiva Filosófica", do PPGFIL/UFPE, gostaríamos de dar boas vindas a este número especial "Filosofia da Lógica".

Este número marca a força da tradição em Filosofia e História da Lógica e de Lógica Filosófica no Brasil ao oferecer para a comunidade a consolidação do trabalho de um grupo importante de jovens pesquisadores. Aspiramos que os 26 artigos aqui publicados possam fomentar uma rede de pesquisadores interessados em filosofia e em lógica para avançar em novas agendas de diálogo acadêmico em alto nível em nosso país.

A "Perspectiva" continua assim cumprindo o seu papel e compromisso acadêmico de publicar trabalhos relevantes e encorajar o desenvolvimento da pesquisa qualificada em Filosofia no Nordeste e no Brasil.

Que este número possa trazer discussões instigantes em momento tão difícil no cenário nacional.

Aproveitamos para destacar o excelente trabalho desempenhado pelos representantes discentes na edição deste (longo) volume especial. Henrique Capeleiro e Hugo Mota, muito obrigado!

Marcos Silva
Loraine Oliveira
Editores-chefe

APRESENTAÇÃO DO NÚMERO ESPECIAL SOBRE FILOSOFIA DA LÓGICA

A área de lógica tem uma longa e profícua história na academia brasileira. Abrangendo áreas como filosofia, matemática e computação, a comunidade de lógica tem sido bem-sucedida em produzir pesquisas relevantes.

Neste número especial da Revista Perspectiva Filosófica, uma revista de acesso livre e gratuito vinculada à Universidade Federal de Pernambuco, oferecemos para a nossa comunidade um panorama da pesquisa em filosofia da lógica no atual cenário nacional com jovens e talentosos pesquisadores brasileiros e/ou realizando pesquisas no Brasil.

Ficamos orgulhosos em apresentar para a comunidade 26 artigos originais de uma nova geração de pesquisadores. Estes artigos tratam da filosofia da lógica, em sentido amplo, e foram selecionados pela qualidade, originalidade e representatividade. Assuntos importantes foram abordados, como a filosofia das lógicas não-clássica, aspectos diversos da filosofia da história, a relação da lógica formal com a lógica informal, o anti-excepcionalismo, a revisão e normatividade da lógica, o pluralismo lógico, a natureza da prova e das contradições, a disputa entre realismo e antirrealismo, entre teoria da prova e teoria dos modelos, alguns aspectos filosóficos do intuicionismo, paraconsistência e do quadrado das oposições e ainda aspectos lógicos e filosóficos relevantes da filosofia da matemática.

Esperamos que este número seja representativo dos temas e problemas filosóficos concernentes à lógica e dos problemas lógicos concernentes à filosofia pelos quais se interessam esta nova geração de pesquisadores.

Evandro Gomes abre o nosso número especial, elucidando um pseudo-dilema do conflito entre a filosofia e a lógica. Alguns autores defendem a lógica como parte constitutiva da filosofia. Outros as consideram apenas mero instrumento. Evandro trata especificamente do cenário filosófico brasileiro defendendo que a lógica é indispensável a qualquer boa filosofia.

Nastassja Pugliese, em sua contribuição, argumenta que há uma continuidade entre os métodos formais e informais de demonstração na lógica. Pa-

ra mostrar esta continuidade, ela parte da noção fundamental de inferência lógica para estabelecer os critérios de diferenciação entre o formal e o in-formal, mostrando como a história filosófica da lógica se transforma e como lógica formal e informal podem ser beneficiadas mutuamente.

O artigo de César Frederico dos Santos considera a questão sobre a natureza descritiva e normativa da lógica à luz de resultados da psicologia do pensamento e da psicologia do desenvolvimento. Já Diogo Dias, em sua contribuição, mostra que a lógica aristotélica continua sendo incompatível com a lógica clássica mesmo que Aristóteles reconhecesse a possibilidade de proposições categóricas com termos vazios. Dias defende que isto não afeta a consistência interna da lógica aristotélica.

O artigo de Alexandre Costa-Leite desenvolve a ideia de se definir oposições lógicas em termos de um segmento de linha e propõe um método definindo-as usando um objeto sem dimensões como o ponto. Já Gisele Secco trata, em seu artigo, de práticas de ensino de filosofia da matemática e propõe uma comparação entre os principais traços da noção leibniziana de conhecimento simbólico e algumas passagens tractarianas sobre a aritmética. Isto redunda, segundo ela, em uma nova luz sob as especificidades da definição tractariana de número em comparação com as de Frege e Russell.

Alessandro Bandeira, em seu artigo, discute a lógica de classes de George Boole. Em particular, Bandeira concentra-se nas questões de escolha de axiomas lógicos e de método de solução de equações, argumentando que a escolha dos axiomas em Boole está totalmente associada à álgebra binária na qual as variáveis podem receber valores 0 ou 1. Já Bruno Bentzen, em sua contribuição, revisita a teoria do sentido e da referência de Frege no contexto construtivo da teoria do tipo. Bentzen propõe a identidade computacional como critério para sentidos e argumenta que esta estratégia válida a interpretação do princípio de equipolência de Frege para sentenças e termos singulares. Também tratando de Frege, o artigo de João Vitor Schmidt defende que a definição de ancestral não é circular, não obstante sua inevitável impredicatividade.

Vincenzo Cicarelli, em seu trabalho, também discute o conceito de Lógica de Frege e mostra que o padrão definicional dos princípios de abstração inclui um requerimento adicional que não satisfaz a lei número V de Frege enquanto é satisfeita por abstrações de ordem superior como o princípio de Hume. Já o trabalho de Bruno Ramos Mendonça defende que o problema de Frege e a sua solução em termos da distinção sentido-referência são generalizáveis a uma família mais ampla de casos de falha de onisciência lógica associados à nossa competência semântica no uso de quantificadores.

A contribuição de Hudson Benevides defende que qualquer tentativa de dispensar a identidade em termos de equivalência lógica entre linguagens de primeira ordem não pode ser bem sucedida. Ao passo que Pedro Carné defende a possibilidade de pensarmos singularmente objetos matemáticos. Para isso, após caracterizar tanto os pensamentos singulares, quanto o problema dos pensamentos singulares sobre objetos matemáticos, Carné reconstrói duas teorias que comprometem-se com esta possibilidade.

John Lennon Lindemann, em seu trabalho, analisa as noções de *nonsense* de Lewis Carroll e Wittgenstein, concluindo que ambos tiveram atitudes análogas em relação ao *nonsense*, usando-o como um meio para induzir ideias instrutivas em seus leitores.

Já o artigo de Newton Peron se debruça sobre alguns problemas envolvendo termos singulares em cinco autores: Frege, Russell, Strawson, Kripke e Kaplan. Newton propõe uma combinação de três tipos de lógica, de modo que tal combinação acomodaria grande parte das filosofias da linguagem quanto aos termos singulares destes autores.

Anderson Nakano, em seu trabalho, revisita o clássico problema dos nomes próprios vazios, ou ainda, o problema dos enunciados existenciais singulares negativos. Partindo de um parentesco deste problema com o pro-blema (ou paradoxo) do falso, Nakano articula discussões entre Aristóteles, primeiro Wittgenstein, Kripke e o Wittgenstein do "Sobre a Certeza" para entender a relação de prioridade entre sentido e valor de verdade. Já a contribuição de Edson Bezerra argumenta em favor da com-patibilidade da Tese de Suszko com a tese do Pluralismo Lógico, e que isto implica certa independência dos

valores lógicos verdadeiro e falso em relação aos Princípios de Não-Contradição, Terceiro Excluído e Explosão.

Hermógenes Oliveira argumenta, em seu ensaio, que uma combinação das posições do pluralismo e do inferencialismo resulta numa postura coerente, robusta e proveitosa para se pensar a filosofia da lógica. Por outro aldo, em seu artigo, Tiago Castro Alves oferece uma caracterização da questão da identidade de provas como um problema concernente à semântica de argumentos dedutivos, pertinente à lógica e à filosofia da linguagem. Alves nos oferece uma descrição dos principais aspectos formais discutidos quanto a suas possíveis motivações e consequências relevantes.

Neste contexto inferencialista, Ralph Heck avalia uma formalização do jogo de dar e pedir razões de Brandom que o divide em duas instâncias, uma individual de performances potencialmente públicas e outra efetivamente histórico-pública, e propõe correções à primeira das duas instâncias. Heck trata da revisão de modelagens adicionando elementos pragmáticos às operações sobre crenças.

Evelyn Erickson, em seu artigo, discute o anti-excepcionalismo lógico atribuído a Quine, assim como duas propostas mais recentes para o método de revisão de teorias lógicas, a saber, a abdução e o equilíbrio reflexivo. Já, Sanderson Molick mostra que, dentre as visões usuais do conhecimento matemático, as mais conhecidas tentam erguer uma base para o conhecimento matemático a partir de elementos *a priori*, como a noção de prova, ou de “intuição primordial”. Molick, em seu trabalho, tem como objetivo apresentar uma visão antagônica às posições apriorísticas, a saber, a do anti-excepcionalismo lógico.

Maria Martinez-Ordaz e Josafat Hernandez defendem, em seu artigo, que as contradições entre teoria e observação e as práticas de tolerância de contradições em ciências sociais oferecem um terreno fértil para o estudo do raciocínio inconsistente, mas não-trivial. Já, o artigo de Ederson Melo e Jonas Arendhart argumenta que o procedimento de "esticar a verdade", permitindo aglutinações, garante a paraconsistência ao custo de distorcer a interpretação da noção de contradição envolvida no dialeteísmo. Especificamente, eles de-

fendem, que tal distorção enfraqueceria a interpretação da negação comprometendo o sentido de contradição relevante para o dialeteísmo.

Por outro lado, Abílio Rodrigues e Henrique Antunes, em sua contribuição, apresentam duas abordagens não dialeteístas da paracon-sistência, discutindo as principais motivações e os conceitos básicos de algumas lógicas paraconsistentes. Em ambas alternativas, eles argumentam que não é necessário endossar a tese dialeteísta de que a realidade é intrinsecamente contraditória para justificar a legitimidade das lógicas para-consistentes.

Fechamos este número especial com a contribuição de Guilherme Cardoso, Ederson Safra e Túlio Aguiar onde discutem aspectos técnicos e filosóficos de uma família de versões do Paradoxo do Mentiroso, mostrando que é impossível se conciliar completamente poder expressivo ilimitado de um lado e consistência ou não-trivialidade de outro.

Nós aspiramos que este número especial possa encorajar novas agendas de pesquisa ao fomentar colaborações acadêmicas no Brasil lançando luz seminalmente para problemas filosóficos da lógica e problemas lógicos da filosofia.

Marcos Silva

Francicleber Ferreira

Editores

UMA HIPÓTESE SOBRE UMA FALSA DICOTOMIA ENTRE LÓGICA E FILOSOFIA

Evandro Luís Gomes[1]

RESUMO

A fim de elucidar uma falsa dicotomia entre a filosofia e a lógica, enunciamos uma hipótese que credita a gênese de tal tensão à índole das relações (efetivas ou pretendidas) entre lógica e filosofia. Da relutância ao desabono, duas posturas teóricas parecem explicar o presente cenário, *i.e.*, se um determinado autor ou escola filosófica considera ou não a lógica parte constitutiva da filosofia ou se seria dessa apenas mero instrumento. Procuramos explicar as razões dessa indiferença e reiteramos, por fim, como a lógica é indispensável a qualquer boa filosofia.

Palavras-chave: Lógica. Dialética. Órganon. Filosofia. Ensino de Lógica. Filosofia Moderna.

ABSTRACT

In order to elucidate the false dichotomy between philosophy and logic, I analyze the hypothesis that credits the genesis of such concern to the nature of the relations (effective or intended) between logic and philosophy. From reluctance to despondency, two theoretical positions seem to explain the present scenario, *i.e.*, as whether a particular author or philosophical school considers or not logic a constitutive part of philosophy, or whether it is just its mere instrument. Since Brazilian philosophical scenario expresses a certain reticence and, at times, disregard towards logic, I proceed to explain the reasons for this indifference and so reiterate, finally, how logic is indispensable to any good philosophy.

Keywords: Logic. Dialectic. Organon. Philosophy. Logic Teaching. Modern Philosophy.

[1] Departamento de Filosofia, Programa de Pós-Graduação em Filosofia, Centro de Ciências Humanas, Letras e Artes, Universidade Estadual de Maringá (PR). Agradecemos aos organizadores do I Colóquio de Lógica da Universidade Federal de Uberlândia, especialmente aos professores Márcio Chaves-Tannus e Marcos Seneda, pelo convite para a conferência que originou o presente ensaio. Agradeço também aos professores Itala M. Loffredo D'Ottaviano, Ricardo Pereira Tassinari e Fábio M. Bertato, pelas frutíferas discussões acerca do tema. E-mail: elgomes@uem.br. ORCID: 0000-0002-7570-6791.

Introdução

As origens da lógica no pensamento ocidental remontam aos primórdios da filosofia grega. Gestada naquele contexto, foi só aos poucos que a lógica – então designada por diferentes nomes, sendo analítica[2] e dialética[3] os mais célebres – foi distinguida e constituída em campo próprio de estudos.[4]

Mesmo considerando as nuances de escola, métodos e teorias, é possível identificar uma constante quanto ao objeto a que visa a lógica através de sua história. Pode-se afirmar, de modo geral, que a lógica se dedica ao estudo dos princípios, leis e regras concernentes à inferência válida ou correta. Seu propósito, explica Bocheński (1966, p. 2, 40), consiste em determinar as condições formais da validade lógica, bem como a construção de seus respectivos modelos, seja no escrutínio da noção de demonstração, seja no exame da noção de dedução formal tal como empregada na análise de argumentos. Assim concebida, assere Bocheński (1951, p. 9, 14, 19; 1966, p. 110), a lógica constituiu-se apenas a partir do legado de Aristóteles, pois foi somente com o Estagirita que ela se tornou disciplina autônoma e seu objeto passou a ser estudado por sua própria sorte.

Se ainda hoje persiste alguma tensão entre lógica e filosofia na percepção de alguns acadêmicos, neste ensaio, analisamos uma hipótese que credita à índole das relações (efetivas ou pretendidas) entre lógica e filosofia a eluci-

---

[2] Tal denominação vincula-se à tradição peripatética. Na verdade, explica Bocheński (1966, p. 57), "Aristóteles no posee ninguna denominación técnica para la Lógica: lo que nosotros denominamos hoy día 'lógico', en él recibe el nombre de 'analítico' (ἀναλυτικός) o 'que se sigue de las premisas' (ἐκ τῶν κειμένων), mientras la expresión 'lógico' (λογικός) significa lo mismo que nuestro 'probable' o bien 'epistemológico'".

[3] Essa denominação está associada à tradição megárico-estoica. De fato, argumenta Blanché (2001, p. 93), "a expressão consagrada de 'lógica dos estoicos', que não teremos escrúpulos em empregar, é, com todo o rigor, duplamente inexata. Primeiro, porque o que nós chamamos a sua lógica, os estoicos chamavam-lhe antes dialética, designando neles (sic) a palavra 'lógica', num sentido mais amplo, tudo o que se refere à linguagem, λόγος, incluindo a retórica e a gramática". Em segundo lugar, sustenta Blanché (*Id., ib.*), "por muito reduzidas que sejam as informações que possuímos sobre os megáricos, o que deles sabemos basta para nos certificar que são os verdadeiros fundadores da lógica dita estoica, e seria de toda justiça qualificar de preferência tal lógica de megárico-estoica". Não obstante, perante o fato de que a forma madura da lógica estoica foi alcançada apenas com Crisipo de Soli, a posição de Blanché, que segue a de Bocheński (1966, p. 117), levanta certa controvérsia entre os estudiosos.

[4] É com propriedade que Guthrie (1962, vol. 1, p. xi; nossa tradução) afirma que "muito do que poderia agora ser considerado como filosófico – teoria ética e política, lógica e epistemologia – ou é totalmente ausente nesse período [pré-socrático], ou se apresenta num estágio embrionário".

dação de tal inquietação. Chegamos, assim, ao cerne do que denominamos "falsa dicotomia entre lógica e filosofia", que tem se traduzido, por vezes, em relutância filosófica à lógica tanto em certos momentos da história da filosofia quanto em certos círculos filosóficos do tempo presente.[5] A fim de compreender as razões dessa tensão epistemológica que, por vezes, se converte em menoscabo, procuramos descrever a hipótese de que a relação entre lógica e filosofia, bem como os papéis esperados e desempenhados pela lógica em diferentes contextos teóricos, está na base dessa atitude. Assim, examinamos duas posturas teóricas, *i.e.*, se um determinado autor ou escola filosófica considera ou não a lógica (*i*) mero instrumento da filosofia ou (*ii*) seria dela parte constitutiva.[6]

1. Sobre as relações entre lógica e filosofia

Enquanto saber racional, parece pacífico que a filosofia não possa prescindir da lógica, pois esta, argumenta Newton da Costa (1980, p. 4), em seus diferentes níveis de análise, formalismo e linguagem, codifica o caráter inferencial da atividade racional e teórica. Não por acaso, já entre os filósofos pré-socráticos encontram-se pistas valiosas que registram o nascimento da lógica entre os gregos. Foi a partir da análise minuciosa do discurso ordinário, da necessidade premente de distinguir as consequências logicamente lícitas das ilícitas, que o desenvolvimento da lógica pôs-se em movimento. Primeiro, como explica Santos (2008, p. 1-30), no contexto da isegoria típica da pólis grega, no qual a lógica emerge do exercício argumentativo político-jurídico, e, paulatinamente, com o incremento do debate filosófico, a lógica dirige-se à argumentação teórica em crescente contraste com a sofística, a fim de caracterizar efetivamente a correlação entre verdade e demonstração. A primeira contribuição notável, nesse sentido, é-nos legada por Parmênides. Os desdo-

[5] No que tange aos "círculos filosóficos do tempo", a presente percepção é pessoal, porém retrata vivências acadêmicas análogas de diversos atores da comunidade lógico-filosófica.

[6] Há outros aspectos que pertencem ao problema em tela, como o contraste entre a eficácia teórico-dedutiva e as expectativas epistêmicas cumuladas à lógica, assim como aqueles que se derivam do magistério de lógica em determinados períodos da história e às aspirações de modernidade em confronto com tradições lógico-filosóficas historicamente estabelecidas. Todavia, pretendemos encetar tal discussão, deveras interessante, em outra oportunidade.

bramentos ontológicos e lógico-epistêmicos das teses parmenidianas sobre aquilo que é (τὸ ὁν), as quais configuram proto-enunciados dos princípios de identidade, do terceiro excluído e de não contradição – e acerca das vias do conhecimento –, são, assim, decisivos para toda a filosofia ocidental.[7]

1.1 A lógica enquanto instrumental para a filosofia

Apesar de a lógica propriamente formal e de caráter nomológico ter se estabelecido enquanto disciplina somente a partir da contribuição aristotélica, sabe-se que esquemas de inferência e princípios lógicos foram utilizados e enunciados no período pré-aristotélico da história da lógica grega (Bocheński, 1951, p. 14-18). Zenão de Eleia, nesse sentido, desempenha um papel mais importante do que se costuma reconhecer no tocante à história da lógica ocidental.

Embora Zenão frequentemente seja encarado como feroz defensor do monismo ontológico de Parmênides, atualmente essa interpretação é bastante disputada, sendo possível sustentar que ele argumenta indistintamente contra ambos, tanto o pluralismo quanto o monismo.[8] Desse modo, seu intento parece ser dialetizar as diversas posições filosóficas em debate à época.

A posição de destaque que Zenão ocupa no panorama filosófico helênico deve-se, sobretudo, às conclusões deconcertantes a que chegou por meio de esquemas de inferência apagógicos.[9] Zenão, pela primeira vez no Ocidente, aponta para a existência da distinção entre forma e conteúdo da argumentação no debate filosófico. Provavelmente essa é a razão de ele ter sido considerado por Aristóteles o fundador da dialética, conforme relata Diógenes Laércio: "Aristóteles diz no *Sofista* que Empédocles foi o inventor da retórica, e Zenão, da dialética".[10] Considerado fidedigno, esse relato circunscreve-se a um con-

[7] Austin (1986, p. 9). A propósito dessa leitura, alguns estudiosos defendem que a Parmênides, e não a Aristóteles, deveria ser outorgado o título de fundador da lógica; *vide* Enriques (1948, p. 8-10).

[8] Cf. Barnes (2000, p. 235-236) e Cordero (1994, vol. II, p. 17-21). Barnes o considera mais um autor perspicaz que um devoto defensor das ideias de Parmênides.

[9] São apagógicas as estratégias de demonstração indireta e de caráter refutativo, como o método de redução ao impossível ou absurdo. Para uma discussão mais detalhada dos métodos dialético-apagógicos de Zenão, *vide* Gomes & D'Ottaviano (2017, p. 70-75).

[10] *Vitae philosophorum* VIII, 57. Tradução de M. da G. Kury, cf. Diógenes Laércio (2008, p. 241). DK29A10: "Ἀριστοτέλης δ' ἐν τῷ Σοφιστῇ φησι πρῶτον Ἐμπεδοκλέα ῥήτορικὴν

texto específico, em que Aristóteles e os doxógrafos teriam reconhecido semelhanças entre o método de Zenão e o dos filósofos megáricos, tanto exercido por Sócrates quanto exemplificado nos primeiros diálogos de Platão.[11] Daí a dialética zenoniana constituir-se, segundo Cordero (1994, p. 21; 33), em uma "técnica de discussão", em especial como exercício de inferência, uma espécie de ginástica mental, um método sem qualquer orientação filosófica específica. De acordo com Cordero, o relato de Simplício testemunha essa interpretação: "Parece que Zenón intentaba, a modo de ejercitación (*gymnastikós*), demonstrar ambas cosas (o sea, lo uno y la multiplicidad). Por eso era llamado 'bilingüe'; y planteaba dificultades acerca de lo uno con estos argumentos" (*In Physica* 139, 3-6, *apud* Cordero, 1994, p. 32, 36). Zenão parece, assim, ter sido o primeiro a utilizar-se de esquemas formais de demonstração e a empregá-los deliberadamente na investigação filosófica.

Outra novidade metodológica de Zenão de Eleia é a consideração em separado da forma com relação ao conteúdo da argumentação desenvolvida. E, de acordo com Cordero (1994, p. 20), nisso consiste sua originalidade filosófica, *i.e.*, o emprego puramente formal de esquemas de argumentação em dialética. Precursor dos *dissoi logoi* de Protágoras, Zenão apresenta uma contribuição à história da filosofia e da lógica que parece ter sido de índole metodológica. Assim, explica Cordero (1994, p. 20-21), o filósofo teria tornado independentes de seus contextos conceituais certos métodos formais, como a redução ao absurdo, a fim de argumentar a favor e contra determinadas hipóteses. Tal acepção "instrumental" da lógica se vislumbra também em Platão.[12] Embora a dialética seja muito cara ao mestre de Aristóteles, não é no sentido de uma disciplina própria, de índole formal, que ela ali se efetiva. De acordo com Bocheński (1966, p. 45), mesmo tendo Platão levado em conta esquemas de inferência lógicos e nitidamente empregados em toda a sua obra, e ainda

εὑρεῖν, Ζήνωνα δὲ διαλεκτικήν". O *Sofista* seria um diálogo aristotélico perdido que não chegou até nós.

[11] Cf. Cordero (1994, p. 20-21); *vide* também Kirk, Raven & Schofield (1994, p. 290).

[12] De acordo com Laudan (2000, p. 27-28), provém do relato de comentadores antigos como Simplício, que repercute a opinião de Geminus (primeiro século depois de Cristo), de que Platão estaria, *mutatis mutandis*, igualmente vinculado à tradição instrumental em ciência, ao ser reconhecido como fundador da tradição grega posterior de "salvar os fenômenos" (*apparentias salvare*).

que tenha enunciado pela primeira vez um conceito correto de lógica no *Timeu* (47 B-C), ele nunca estudou a lógica por si mesma.

Embora Zenão pareça ter sido o primeiro a conceber a lógica (dialética) como instrumento do filosofar, essa concepção de lógica foi vinculada a Aristóteles e à sua escola, como consagra a tradição posterior.[13] Assim sendo, apesar de Aristóteles empregar o termo organon (ὄργανον) em diversas passagens (cf. *Top.* 108b 30; 163b 10), a imposição à lógica de um caráter instrumental, com fundamento em alguns acenos no *corpus aristotelicum*, deve-se, sobretudo, aos comentadores gregos do Estagirita.[14]

De fato, Aristóteles não teria deixado tão clara sua posição quanto à lógica ser parte da filosofia ou dela mero instrumento. Com efeito, argumenta Bocheński (1966, p. 57, nossa tradução), "a questão de se a lógica é uma parte da filosofia ou seu instrumento (ὄργανον) e, por conseguinte, uma arte, não tratou dela Aristóteles em nenhuma parte dos escritos que conservamos". Entretanto, em um das poucas passagens em que aborda o tema abertamente, Aristóteles indica que em toda arte (τέχνη) são necessários instrumentos e que esses são de muitas espécies (cf. *Pol.* 1253b 28). Todavia, em nenhum lugar ele afirma que a lógica é um instrumento da filosofia. Não obstante, encontra-se em Aristóteles a defesa de que a lógica é não apenas uma parte da filosofia, mas uma disciplina preparatória, um pré-requisito ao filosofar:

> Quanto a tudo aquilo que alguns (entre os que foram mencionados) tentam estabelecer a respeito da verdade, concernente ao modo pelo qual é preciso aceitá-la, fazem-no devido à falta de formação nos Analíticos; de fato, é preciso chegar já sabendo previamente esses assuntos, mas não buscá-los enquanto se ouve [*sc.* o presente curso].

---

[13] Com efeito, Andrônico de Rodes, quando resgatou os escritos de Aristóteles, por volta do I século a.C., agrupou os tratados lógicos como hoje conhecemos, porém a denominação "órganon" é ainda posterior; *vide* Bocheński (1966, p. 52-53) e Blanché (2001, p. 27-28, n. 1, 2 e 3). Kneale & Kneale (1991, p. 25) explicam que "quando os escritos de Aristóteles foram reunidos pelos seus alunos depois da sua morte em 322 a.C., uma série dos seus tratados sobre o raciocínio foi agrupada e a coleção acabou por se chamar *Organon*, ou instrumento da ciência. A palavra 'lógica' só adquiriu o seu sentido moderno 500 anos mais tarde quando foi usada por Alexandre de Afrodísias; mas o âmbito da investigação, mais tarde chamada lógica, foi determinado pelo conteúdo do *Organon*".

[14] De fato, Alexandre de Afrodísia e João Filopono que teriam difundido tal uso; cf. Ferrater Mora (2000, t. 3, p. 2170-2171). Tais comentadores, entretanto, não se referiam a todos os tratados indistintamente, conjuntamente considerados, mas apenas ao seu conteúdo, à doutrina lógica ali exposta.

> Assim, é evidente que compete ao filósofo, isto é, àquele que estuda como naturalmente se apresenta qualquer essência, investigar também a respeito dos princípios silogísticos. E convém que aquele que mais conhece a respeito de cada gênero seja capaz de enunciar os princípios mais firmes do assunto, de modo que também aquele que conhece a respeito dos entes enquanto são entes é capaz de enunciar os princípios mais firmes de todas as coisas. E este é o filósofo. (*Met.* Γ 1005b)[15]

Consequentemente, carece de fundamento a opinião amplamente difundida de que a lógica seria, para os peripatéticos, um mero instrumento, e não parte da filosofia. Essa opinião reforça a leitura de que, para Aristóteles, seria forçada a separação entre lógica e filosofia. Ademais, isso implica, precisamente, que a lógica ali só poderia ser arbitrariamente separada da filosofia. Nesse sentido, explica Correia (2002, p. 11, nossa tradução),

> a chamada "lógica de Aristóteles" é uma consequência de uma teoria dialética e ontológica que, se bem possa converter-se desde este mesmo princípio numa disciplina formal, não pode identificar-se somente com ela. Daí que a lógica, para seu fundador, não é nem uma dialética, nem uma ontologia, nem uma lógica formal, senão tudo isso junto, historicamente fundido.

Desse modo, foi a partir da indicação aristotélica do caráter propedêutico da lógica (*i.e.*, da analítica) que se amoldou e repercutiu a opinião dos comentadores antigos, de que a lógica fosse encarada como instrumento da filosofia e das ciências. A tradição lógica e filosófica medieval teria aprofundado a indicação dos comentadores de que a lógica seria instrumento da filosofia. Parece aí haver uma especial influência da revalorização aristotélica ocorrida na Antiguidade tardia e na filosofia escolástica, o que teria levado, inclusive, à pretensão de um purismo aristotélico. Um exemplo dessa concepção encontra-se na Carta-Prefácio da *Suma de toda a lógica*. Guilherme de Ockham considera a lógica essencial para o desenvolvimento de todas as demais ciências e exemplifica a ideia da lógica como uma arte, mas também como instrumento imprescindível ao empreendimento da atividade racional:

> Pois a lógica é a mais útil de todas as artes. Sem ela, nenhuma ciência poderia ser totalmente conhecida. Ela não se esgota pelo uso repetido, segundo a maneira das ferramentas materiais, mas antes admite crescimento contínuo por meio do exercício diligente de qual-

[15] Tradução de Lucas Angioni; cf. Aristóteles (2001, p. 18).

> quer outra ciência. Pois exatamente como um mecânico, a quem falta um completo conhecimento de sua ferramenta, ganha um mais completo [conhecimento] por usá-la, assim alguém que é educado nos firmes princípios da lógica, enquanto meticulosamente devota seu trabalho a outras ciências, adquire ao mesmo tempo uma habilidade maior nessa arte. (*Summa totius logicae*, Epistola Proemialis, nossa tradução)[16]

Semelhante elogio à lógica jamais poderia desaboná-la. Ao longo da história da filosofia e no compasso do alvorecer da matematização da lógica, sobretudo no século XIX, muitos estudiosos que não assimilaram essa radical transformação no cerne da disciplina passaram e vê-la como algo estranho ao seu próprio ofício, ao mesmo tempo que se sentiam capazes de desenvolver seus temas e problemas sem recorrer aos métodos da lógica de suas respectivas épocas, em especial, mais recentemente, com sua apresentação simbólica e matematizada.[17] Eis uma das facetas da falsa dicotomia entre lógica e filosofia. Por outro lado, ao encararem seus próprios interesses teóricos em filosofia como lógico-independentes, muitos estudiosos parecem considerar suficiente a posse de um pequeno arsenal metalógico de índole informal e de natureza lógico-clássica. Métodos tradicionais em história da filosofia, tal como consagrada no Brasil – de índole hermenêutica e estrutural –, fundamentam a análise dos temas e autores a partir de seus conceitos-chave e de maneira interna a seus próprios sistemas de pensamento, como se a arquitetura dos conceitos entrelaçados por esses autores pudesse oferecer-lhes as ferramentas necessárias. Ou seja, existe uma suposição de que a lógica interna ao discurso de um autor basta à correta discussão de um sistema de conceitos. Aqui, a lógica de que falamos desde o princípio é preterida em favor de uma racionalidade localíssima, uma "lógica", sem dúvida, muito peculiar; trata-se, como a entende-

---

[16] "*Logica enim est omnium artium aptissimum instrumentum, sine qua nulla scientia perfecte sciri potest, quae non more materialium instrumentorum usu crebro consumitur, sed per cuiuslibet alterius scientiae studiosum exercitium continuum recipit incrementum. Sicut enim mecanicus sui instrumenti perfecta carens notitia utendo eodem recipit pleniorem, sic in solidis logicae principiis eruditus dum, aliis scientiis operam impendit sollicite simul istius artis maiorem adquirit peritiam. Unde illud uulgare 'ars logica labilis ars est' in solis sapientiale studium negligentibus locum reputo obtinere.*"

[17] É conhecido o desabono dos métodos da lógica pelos filósofos modernos, seus herdeiros e intérpretes. Embora haja honrosas exceções, como Leibniz e Kant, é pacífico afirmar que a lógica não atendia às aspirações epistemológicas daqueles filósofos que nela procuraram – sem, contudo, encontrar –, sobretudo, uma arte da descoberta (*ars inveniendi*). Cf. Cassirer (1994).

mos, de um aprofundamento antitético da visão de que a lógica, enquanto disciplina historicamente estabelecida e teoricamente fundante, não é nem sequer instrumento da filosofia.

1.2 A lógica enquanto parte intrínseca da filosofia

Um exemplo emblemático do vínculo da lógica à filosofia é-nos legado pelos estoicos. Sua teoria lógica, como parte de um todo sistemático, articula-se com as outras partes de sua filosofia, particularmente com a sua teoria do conhecimento. Nesse aspecto, há um contraste notável entre a postura que se origina da tradição peripatética, que concebe a lógica de forma instrumental, uma verdadeira propedêutica filosófica, e a atitude estoica de integrar a lógica sistemicamente à sua filosofia.[18] São-nos legadas por Diógenes Laércio as belas e inconfundíveis metáforas com as quais a lógica é descrita como parte constitutiva da filosofia:

> Os estoicos comparam a filosofia a um ser vivo, onde os ossos e os nervos correspondem à lógica, as partes carnosas à ética e a alma à física. Ou então comparam-na a um ovo: a casca à lógica, a parte seguinte (a clara) à ética, e a parte central (a gema) à física. Ou a comparam ainda a um campo fértil: a cerca externa é a lógica, os frutos são a ética, e o solo ou as árvores são a física. Ou comparam-na a uma cidade bem amuralhada e racionalmente administrada. E nenhuma parte é separada das outras, como dizem alguns estoicos, mas ao contrário todas estão estreitamente unidas entre si. (*Vitae philosophorum*, VII, 40)

Faz-se mister recordar que λογική (*logiké*) designava, no mundo antigo, tudo aquilo que se situava no âmbito da razão ou do lógos. Assim, embora o termo "lógica" seja uma transliteração do termo grego, seus referentes atual e original não coincidem exatamente. Para os estoicos, a lógica abrangia a retórica e a dialética. O sentido contemporâneo de lógica seria, em parte, compatível, como antecipamos, com aquilo que os estoicos denominavam διαλεκτική (*dialektiké*). Ainda assim, a dialética antiga tinha um sentido mais amplo que o da lógica atual; a dialética estoica, por exemplo, abrangia os significados e o som vocal. O âmbito dos significados subdividia-se no estudo das "representações", como o som e a voz, e naquilo que subsistia a partir delas:

[18] *Vide* também Blanché (2001, p. 108).

os *lekta*.[19] Assim, os dizíveis, ao traduzirem as representações (φαντασίαι), colimam a dialética estoica à sua contraparte epistemológica.[20]

É oportuno ressaltar que a lógica estoica, primorosamente sistematizada por Crisipo de Soli, alcança, na Antinguidade, grande reputação, maior até que aquela então obtida pela lógica peripatética,[21] donde se depreende que a visão segundo a qual a lógica constitui parte intrínseca da filosofia foi igualmente bem difundida. Uma passagem de Clemente de Alexandria, autor cristão do final do segundo e do princípio do terceiro século da era cristã, dá a dimensão desse prestígio. Ele afirma no *Stromata* (VII, 16, 323) que Crisipo é o mestre dentre os lógicos, Aristóteles é o mestre dentre os cientistas, Platão é o mestre dentre os filósofos e Homero o é dentre os poetas. A escola estoica mantinha grande vigor ainda no tempo de Marco Aurélio (†180).[22]

É bem conhecido que a ética constituía o propósito máximo de toda reflexão filosófica no período helenístico. Nesse sentido, quanto ao estoicismo, salienta Barnes (1997, p. 10-11; nossa tradução): "A física e a lógica eram abordadas apenas na medida em que poderiam servir a um fim ético, e o objetivo e o propósito da filosofia era apresentar ao homem como melhor conduzir a sua vida". Nesse contexto, contudo, a atitude quanto à lógica adotada por alguns autores como Marco Aurélio é claramente negativa. Todavia, segundo

---

[19] Diógenes Laércio e Sexto Empírico apresentam os λεκτα como aquilo que subsiste após uma representação lógica. De acordo com Mates (1961, p. 11), um λεκτόν poderia ser traduzido literalmente como "aquilo que é significado". Ildefonse (2007, p. 102-105) sugere que o termo "λεκτόν", apesar de seu complexo significado, corresponde ao sentido de uma frase completa, podendo ser traduzido por "dito". Desse modo, apenas um λεκτόν incompleto corresponderia ao sentido de um termo, de um predicado. Os *lekta* são, juntamente com o tempo, o lugar e o vazio, um dos quatro incorpóreos admitidos pelos estoicos. Ao contrário dos *lekta*, relata Sexto Empírico, os signos (τὸ σημαῖνον) são objetos materiais, assim como todos os demais objetos segundo a filosofia estoica; *vide* Mates (1961, p. 13).

[20] Cf. Ildefonse (2007, p. 100) e Barnes (1997, p. 8-9). Segundo Ildefonse (*Id.*, 102), "o que garante a passagem do evidente da percepção ao não evidente e define o processo da ciência é a representação lógica (*phantasia logiké*) suscetível de composição e de transição (*synthetikè kai metabatikè*), e o logos que desenvolve essa representação". Sobre esse ponto, *vide* Sexto Empírico, *Adv. Math.* VIII, 70 e Diógenes Laércio, *VP*, VII, 63. O último explica que a representação é o efeito sofrido na alma (*pa thos en tèi psychèin*), precisamente em sua parte diretora (*hegemonikon*), e é nessa porção que se produzem os impulsos e as representações, bem como a partir daí é emitida a linguagem; *vide VP*, VIII, 159.

[21] Sobre o declínio da lógica peripatética na Antiguidade tardia, *vide* Speca (2001, p. 3-6).

[22] Para uma análise e interpretação do papel da lógica estoica no estabelecimento de uma abordagem de cunho lógico relevante e paraconsistente na lógica grega antiga, *vide* Gomes & D'Ottaviano (2017, p. 127-142).

Barnes (1997, p. 4), a ideia de que a lógica foi completamente abandonada e relegada ao final do Antigo Stoá, após o II século a.C., parece equivocada.

O cenário, não obstante, é bem diverso desse.[23] A propósito, explica Barnes, "os contemporâneos de Epicteto não eram moralistas cegos, ignorando os estudos lógicos que o Antigo Stoá tinha padronizado. Ao contrário, a filosofia – a filosofia estoica – tinha virtualmente se reduzido à lógica". Esse parecer mostra como, no contexto da lógica estoica da era imperial, as coisas se passavam de modo muito diferente da opinião corrente sobre esse período, do que Barnes (1997, p. 37; nossa tradução) conclui: "Os filósofos eles mesmos pensam apenas em seus silogismos. Seus jovens estudantes eram escolados em técnica lógica e não poderiam encontrar melhor conversa à mesa de jantar que inferências hipotéticas e o argumento Mestre".[24]

As *Diatribes* de Epicteto mostram um autor voltado, em primeiro plano, à temática ética. Apesar dessa tendência marcante, sua obra contém numerosas alusões a assuntos que pertencem à lógica. Do começo ao fim dessa obra, a terminologia lógica pode ser encontrada. Ele menciona, por exemplo, os condicionais (συνημμένα) e as disjunções (διεζευγμένα), as proposições compostas (τροπικά) e os argumentos que são válidos (συνακτικοί) ou inválidos (ἀσύνακτικοι). Ao utilizar tal terminologia, explica Barnes (1997, p. 28), Epicteto o faz sem aviso e sem dar-lhe qualquer destaque especial, como alguém treinado que deixa o vocabulário lógico perpassar naturalmente seu discurso.

Epicteto lega-nos, em suas *Diatribes*, na forma de um argumento por redução ao absurdo logicamente clássica, uma das mais belas apologias à lógica de que se tem notícia. Formalmente falando, trata-se de uma *peritropé* semelhante àquela utilizada por Platão no *Teeteto* (171 B-C), mas que visa

[23] Barnes (1997, p. 37; nossa tradução): "Que a lógica excitou e atraiu era um lugar-comum. Assim, se consideramos o que Epicteto e seus companheiros realmente dizem e sugerem acerca da vida intelectual de seu tempo, vemos um retrato completamente diferente daquele pintado pela visão comum sobre o estoicismo imperial".

[24] O Argumento Mestre, Dominador ou Vitorioso (ὁ κυριεύων λόγος) é um argumento que Diodoro Cronos propôs a fim de estabelecer sua própria definição de possibilidade. É de Epicteto a única versão completa que possuímos desse argumento; cf. Kneale & Kneale (1991, p. 121, n. 3).

demonstrar que o estudo da lógica é necessário.[25] A passagem toda é a seguinte:

> QUÃO NECESSÁRIA É A LÓGICA:
> [1] Quando algum dos presentes disse:
> – Persuade-me de que a lógica é útil.
> – Queres, disse Epicteto, que te demonstre isso?
> [2] – Sim!
> – Portanto, é-me preciso selecionar um argumento demonstrativo?
> Quando o outro concordou, <Epicteto disse>:
> – E como saberás se eu te apresentar um sofisma?
> [3] Quando o homem se calou, Epicteto disse:
> – Vês como tu mesmo concordas que a lógica é necessária, já que sem ela não é possível saber se é necessária ou não.[26]

De acordo com Barnes, a expressão ταῦτα no original indica a lógica (τὰ λογικά), designada, inclusive, no próprio título do capítulo. Em síntese, Epicteto argumenta que, para saber se é ou não necessário estudar lógica, estudar lógica necessário o é. O argumento de Epicteto pode ser assim reescrito:

> (1) Se é necessário estudar lógica, então é necessário estudar lógica.
> (2) Se não é necessário estudar lógica, então [ainda assim] é necessário estudar lógica.
> (3) Portanto, é necessário estudar lógica.[27]

Tal esquema da inferência de Epicteto é bastante próximo ao argumento aristotélico acerca da inexorabilidade do filosofar preservado no *Protréptico*. Aqui, como lá – quando Aristóteles demonstra que, mesmo se alguém se negar a filosofar, filosofará de qualquer modo[28] –, Epicteto conclui que, ain-

---

[25] Barnes (1997, p. 31, n. 34) ressalta que tanto o termo περιτροπή quanto o verbo περιτρέπειν não são encontrados em Epicteto.

[26] *Diatribes*, II xxv: "Τῶν παρόντων δέ τινος εἰπόντος Πεῖσόν με, ὅτι τὰ λογικὰ χρήσιμά ἐστιν, Θέλεις, ἔφη, ἀποδείζω σοι τοῦτο; – Ναί. – Οὐκοῦν λόγον μ᾽ ἀποδεικτικὸν διαλεχθῆναι δεῖ; – ῾Ομολογήσαντος δὲ Πόθεν οὖν ἔστη, ἄν σε σοφίσωμαι; – Σιωπήσαντος δὲ τοῦ ἀνθρώπου ῾Ορᾷς, ἔφη, πῶς αὐτὸς ὁμολογεῖς ὅτι ταῦτα ἀναγκαῖά ἔστιν, εἰ χωρὶς αὐτῶν οὐδ᾽ αὐτὸ τοῦτο δύνασαι μαθεῖν, πότερον ἀναγκαῖα ἢ οὐκ ἀναγκαῖά ἐστιν". Texto estabelecido por W. A. Oldfather; *vide* Epicteto (2000, p. 420; 422). Tradução de Aldo Dinucci; *vide* Dinucci (2015, p. 195-196).

[27] Salientamos, contudo, que o esquema de inferência de Epicteto expresso ao modo estoico assim se apresenta:

> (1) Se o primeiro, então o primeiro.
> (2) <u>E, se não o primeiro, então o primeiro.</u>
> (3) Portanto, o primeiro.

[28] *Vide* Fragmento 51 Rose[3] (Elias, *Prolegomena Philosophiae* 3.17 – 23). *In*: Aristóteles (1985, vol. 2, p. 2.416-2.417). Essa versão do argumento aristotélico pode assim ser formalizada:

$$\big((A \to A) \wedge (\sim A \to A)\big) \to A\,.$$

da que não se queira estudar lógica, ter-se-á que a estudar para distinguir sofismas de argumentos válidos, o que implica (para quem pretende propor argumentos) que estudar lógica sobrevém de qualquer maneira. O esquema de inferência subjacente ao argumento de Epicteto, análogo àquele utilizado por Aristóteles no *Protréptico*, é válido na lógica clássica e equivale à *consequentia mirabilis*.[29] Aqui também se aplicariam, *mutatis mutandis*, as condições postuladas por Kneale para que argumentos assim construídos fossem válidos, ou seja, somente quando a proposição em questão fosse necessária.[30] Assim, uma vez que o estudo da lógica se mostra necessário, necessário é, do ponto de vista da apologia de Epicteto, estudar lógica. Portanto, abandonar o estudo da lógica equivale a renunciar ao filosofar.

Tendo analisado alguns elementos da abordagem peripatética e estoica à lógica, pode-se concluir que uma pretensa dicotomia entre lógica e filosofia não se pode abrigar no legado das grandes filosofias que inauguraram a tradição lógico-filosófica ocidental. Lógica e filosofia são disciplinas afins tanto na tradição peripatética quanto na estoica, sendo que, nesta última, a lógica é explicitamente admitida enquanto constituinte sistêmico da própria filosofia.

2 Lógica e filosofia, uma falsa dicotomia

Temos até aqui argumentado acerca do papel fundamental e fundante da lógica na codificação da atividade teórico-racional e na descrição das inferências tanto filosóficas quanto científicas. Em razão dos elementos já apresentados, aduz-se que as tradições peripatética e estoica valorizam a lógica, não obstante uma certa visão instrumental da disciplina pareça ter se descolado e reelaborado posteriormente na história da filosofia, a partir da concepção propedêutica de lógica, originalmente, de índole peripatética.

Semelhante visão instrumental da lógica como ferramenta, técnica ou arte é aprofundada no Renascimento e na Modernidade. Nesse estádio histori-

---

Vide também Bocheński (1966, p. 43-44).

[29] A *consequentia mirabilis* (ou lei de Clavius), em símbolos, assim se denota:

$$(\sim A \rightarrow A) \rightarrow A .$$

Vide discussão adicional em Gomes & D'Ottaviano (2017, p. 82-84).

[30] Kneale (1966, p. 162-166).

co, ao relegarem boa parte da tradição lógica anterior, os modernos introduzem seus próprios reparos à teoria lógica, em grande medida de caráter epistemológico, mas também, e inclusive, de índole psicologista.[31] Essa é a perspectiva consagrada na *La Logique, ou l'Art de penser* de Port-Royal,[32] o mais emblemático dos textos de lógica da época. Nesse período, na medida em que variaram os interesses teóricos – com exceção feita a Leibniz e, em certa medida, a Kant –, diversos autores modernos passaram a ver a lógica como algo estranho e, por vezes, até avesso à própria filosofia. Tal visão ecoa ainda em nosso tempo, em que a filosofia se apresenta como um mosaico de diferentes escolas, problemas, programas, linguagens e métodos.

A falsa dicotomia está, desde então, constituída.[33] É justamente por entender que a lógica é independente da filosofia que muitos autores e intérpretes declaram dela sua independência. Assim, a lógica passa a ser encarada como algo apartado e, por vezes, também dispensável ao exercício filosófico. Radicam aí, ao lado de diferentes concepções acerca do fundamento do conhecimento para racionalistas e empiristas, as correlações que originam posturas de apreço ou desapreço à lógica.

Encontramos na história da filosofia algumas tentativas de solução para a presente "falsa dicotomia". Advém de Pedro Abelardo (1079-1142), célebre lógico, filósofo e teólogo do século XII, um lúcido argumento capaz de desfazer a falsa dicotomia entre lógica e filosofia. Com efeito, Abelardo desenvolve, no princípio da *Logica ingredientibus*, uma defesa de que a lógica é parte indissociável da atividade filosófica:

> Para aqueles que dentre nós que se introduzem na Lógica, falemos previamente um pouco a respeito de sua característica própria, começando pelo gênero a que pertence, isto é, a Filosofia. Ora, Boécio não chama de Filosofia qualquer ciência, mas apenas a que se detém nas coisas mais elevadas; de fato, não chamamos de filósofos a quaisquer pessoas dotadas de conhecimento, mas apenas àquelas cu-

[31] *Vide* Buickerood (1985) e Cassirer (1994).

[32] *Vide* Arnauld & Nicole (1964).

[33] Ressalve-se que há também na Antiguidade posições adversas quanto à lógica. Sexto Empírico, por exemplo, nunca a aprecia como fundamento viável ao conhecimento. O'Toole & Jennings (2004, p. 484, nossa tradução) explicam que "como cético Sexto pretendia desacreditar as visões de todos os assim chamados 'filósofos dogmáticos'. Com respeito aos estoicos, isto teria envolvido, entre outras coisas, mostrar seu sistema lógico como inútil como meio de realizar inferências ou sustentar demonstrações".

> ja inteligência penetra as sutilezas. Boécio distingue três espécies de Filosofia, a saber: a *especulativa*, que se ocupa da natureza das coisas a ser investigada; a *moral*, que se ocupa da dignidade da vida a ser considerada; a *racional*, denominada Lógica pelos gregos, que se ocupa da ordem dos argumentos a serem compostos. (*Logica "Ingredientibus"*, 1)[34]

É claro que Pedro Abelardo introduz seus enunciados num contexto lógico-filosófico historicamente constituído, a *logica vetus*, na escolástica inicial. Isso não restringe o alcance de sua argumentação, pois, tal como ele, falamos de noções gerais e abrangentes de lógica e filosofia. Nesse contexto, o filósofo deixa claro que a lógica é uma espécie intrínseca à filosofia e que possui como propriedade fundamental a "ordem dos argumentos", o que remete ao *modus operandi* formal típico da disciplina. Esse traço nunca se perdeu, apesar de, em diferentes períodos de sua história, a lógica ter sido cultivada em benefício das prioridades típicas de cada conjuntura histórica. Pedro Abelardo, o *Peripateticus Palatinus*, não vê razões para qualquer desconexão entre lógica e filosofia. Tal dicotomia é afastada quando, na verdade, todos os saberes se beneficiam dos recursos dialéticos da lógica:

> Isso porque, de um certo modo, as demais partes dela [da Lógica] se ocupam, na medida em que se servem de seus argumentos, para provar suas próprias questões. Assim, se se coloca uma questão pertinente à investigação da natureza ou da moral, os argumentos são tirados da Lógica. Contra esses, o próprio Boécio diz que *nada impede que a Lógica seja tanto instrumento como parte de algo, assim como a mão o é em relação ao corpo humano*. (*Logica "Ingredientibus"*; grifos nossos)

Pedro Abelardo divisa, assim, tanto a metalógica quanto a lógica aplicada, como facetas distintas da mesma parte da filosofia racional, todas voltadas a analisar filosoficamente as minúcias das questões teóricas e práticas. Além disso, ele introduz um argumento decisivo na polêmica que aqui analisamos, de que à lógica é indiferente se é instrumento ou constituinte da filosofia: há muitos exemplos dessa duplicidade que atestam a mútua pertinência entre as coisas comparadas. Portanto, da mesma forma que a mão é parte e instrumento do corpo humano, assim também a lógica é, simultaneamente,

[34] Tradução de Carlos A. R. do Nascimento; *vide* Pedro Abelardo (2005, p. 39-40).

instrumento e parte da filosofia e de seus diferentes gêneros de análise e questões.

A fina percepção do autor quanto à natureza da análise metalógica antecipa perspectivas típicas do século XX e faz-se nítida no passo seguinte de sua argumentação, no qual a lógica não é menos lógica quando aplicada a si mesma.

> Além disso, a própria Lógica se apresenta muitas vezes como instrumento de si própria, visto que demonstra também uma questão a si pertinente com argumentos seus, como, por exemplo, a seguinte: o homem é uma espécie de animal. Contudo, nem por isso é menos Lógica, por ser instrumento da Lógica. Assim, também não é menos Filosofia por ser instrumento da Filosofia. O próprio Boécio a distingue das duas outras espécies de Filosofia pelo seu fim próprio que consiste em compor argumentações. Pois, embora o estudioso da natureza componha argumentos, não é o estudo da natureza que o institui para tal, mas apenas a Lógica. (*Logica "Ingredientibus"*; grifos nossos)

Pedro Abelardo entende que a lógica exerce papel regulativo da atividade racional, conformando todo e qualquer empreendimento epistêmico. Nesse sentido, ele não deixa de valorizar a lógica como o manancial do método, pois, apesar de sua natureza formal, essencial na composição dos argumentos, é a lógica que orienta o estudioso na descoberta e na formulação das leis da natureza.[35]

Considerações finais

De tudo quanto dissemos, concluímos que a "falsa dicotomia" entre lógica e filosofia não resiste a um exame atento da própria história da filosofia e da lógica, sendo que apenas forçosamente se pode operar essa separação.

Tal aposta, que minora ou despreza o papel da lógica na atividade filosófica, pode estar associada à ênfase em leituras modernistas da história da filosofia, que sobrevalorizam as questões e a abordagem dos modernos, período

[35] Algumas leis lógicas, explica Pedro Abelardo, podem ser mais do que uma simples relação lógica, pois expressam uma lei da natureza (*lex naturae*); *vide* Martin (1987, p. 390). Desse modo, uma proposição maximal no contexto da teoria dos tópicos exprime, segundo Martin, "uma verdadeira restrição ontológica geral sobre a estrutura do mundo. Proposições maximais fornecem, como se o fossem, esquemas de acordo com os quais leis da natureza poderiam ser geradas".

em que, além de ter estacionado seu percurso criativo, a lógica foi preterida mediante novos interesses teóricos. Há que ressaltar que a grande ênfase no estudo da filosofia moderna em cursos universitários de filosofia resulta em perda de perspectiva, semelhante àquela deformação, tão flagrante quanto naturalizada, que confunde as proporções do globo terrestre com a representação do planisfério de Mercator, com os continentes e os oceanos aparecendo distorcidos. De modo análogo, em filosofia, determinadas visões fazem alguns campos de estudo parecerem maiores ou menores, ou mais centrais ou periféricos do que realmente o são.

Enfim, manifestações de desapreço à lógica nos dias atuais só se podem explicar pela falta de conhecimento do que seja o seu papel na própria história da filosofia, bem como do que seja a lógica e de como suas diferentes formas historicamente constituídas (greco-escolástica, tradicional e simbólico-matemática) exprimem a evolução de uma mesma disciplina tão importante para a crítica do conhecimento como para a própria filosofia.

Referências

ARISTÓTELES. *The complete works of Aristotle: the revised Oxford translation.* Edited by J. Barnes. Princeton: Princeton University Press, 1985. 2 vols. (Bollingen series, LXXI)

____. *Metafísica – Livros IV e VI*. Tradução, introdução e notas de L. Angioni. Campinas: Instituto de Filosofia e Ciências Humanas-Unicamp, 2001. (Textos Didáticos, 45)

ARNAULD, A. & NICOLE, P. *La logique ou L'Art de Penser.* Édition critique présentée par Pierre Clair et François Girbal. Paris: Presses Universitaires de France, 1964.

AUSTIN, S. *Parmenides: being, bounds, and logic*. New Haven, London: Yale University Press, 1986.

BARNES, J. *Logic & the imperial stoa*. Leiden, New York, Köln: Brill, 1997.

____. *The Presocratic Philosophers*. London, New York: Routledge, 2000. (The arguments of the philosophers)

BLANCHÉ, R. *História da lógica.* Atualizado por J. Dubucs. Trad. A. P. Ribeiro e P. E. Duarte. Lisboa: Edições 70, 2001.

BOCHEŃSKI, I. M. *Ancient Formal Logic*. Amsterdam: North-Holland, 1951.

____. *Historia de la Lógica Formal.* Tradução de M. Bravo Lozano. Madrid: Gredos, 1966. (Biblioteca Hispánica de Filosofía)

BUICKEROOD, J. G. "The Natural History of the Understanding: Locke and the Rise of Facultative Logic in the Eighteenth Century". *History and Philosophy of Logic*, 6, 1985, p. 157-190.

CASSIRER, E. *A filosofia do Iluminismo.* Tradução de Álvaro Cabral. 2ª ed. Campinas: Editora da Unicamp, 1994.

CORDERO, N. L. "Zenón de Elea". *In*: GARCIA GUAL, C. (ed.). *Los filósofos presocráticos*. Madrid: Gredos, 1994, vol. II, p. 7-63. (Biblioteca Clásica Gredos, 24)

CORREIA, M. *La logica de Aristoteles: lecciones sobre el origen del pensamiento lógico en la antigüedad*. Santiago: Ediciones Universidad Católica de Chile, 2002.

DA COSTA, N. C. A. *Ensaio sobre os fundamentos da lógica*. São Paulo: Edusp-Hucitec, 1980.

DINUCCI, A. "Epicteto: Diatribes I.7 e II.25 – Relação entre a lógica e as ações convenientes". *Revista Clássica*, vol. 29, n. 1, 2015, p. 193-203.

DIÓGENES LAÉRCIO. *Vidas e doutrinas dos filósofos ilustres*. Tradução do grego, introdução e notas de M. da G. Kury. 2ª ed. Brasília: Editora UnB, 2008.

ENRIQUES, F. *Para la historia de la logica.* Buenos Aires: Espasa-Calpe, 1948. (Historia y Filosofia de la Ciencia)

FERRATER MORA, J. *Dicionário de Filosofia.* São Paulo: Loyola, 2000, 4 tomos.

GOMES, E. L. & D'OTTAVIANO, I. M. L. *Para além das Colunas de Hércules, uma história da paraconsistência: de Heráclito a Newton da Costa*. Campinas: Editora da Unicamp, 2017.

GUTHRIE, W. K. C. *A history of Greek philosophy*. Cambridge: Cambridge University Press, 1962. (Vol. 1: *The earlier presocratics and the phytagoreans*)

ILDEFONSE, F. *Os Estóicos I*. Tradução de M. Pinheiro. São Paulo: Estação Liberdade, 2007.

KIRK, G. S., RAVEN, J. E. & SCHOFIELD, M. *Os filósofos pré-socráticos: história crítica com seleção de textos.* Tradução de C. A. L. Fonseca. 4ª ed. Lisboa: Calouste Gulbenkian, 1994.

KNEALE, W. "Aristotle and the *Consequentia Mirabilis*". *The Journal of Hellenic Studies*, vol. 77, n. 1, 1966, p. 162-166.

KNEALE, W. & KNEALE, M. *O desenvolvimento da lógica.* Tradução de M. S. Lourenço. 3ª ed. Lisboa: Calouste Gulbenkian, 1991.

LAUDAN, L. "Teorias do método científico de Platão a Mach: resenha bibliográfica". *Cadernos de História e Filosofia da Ciência*, vol. 10, n. 2, 2000, p. 1-140.

MARTIN, C. J. "Embarrassing arguments and surprising conclusions in the development of theories of the conditional in the twelfth century". *In*: JOLIVET, J. & DE LIBERA, A. (eds.). *Gilbert de Poitiers et ses contemporains.* Naples: Bibliopolis, 1987, p. 377-400.

MATES, B. *Stoic Logic*. 2ª ed. Berkeley, Los Angeles: University of California Press, 1961.

O'TOOLE, R. O. & JENNINGS, R. E. "The Megarians and the Stoics". *In*: GABBAY, D. M. & WOODS, J. (eds.). *Handbook of the history of logic; Greek, Indian and Arabic Logic*. Amsterdam: Elsevier, 2004, p. 397-522. (Handbook of the History of Logic, 1)

PEDRO ABELARDO. *Lógica para principiantes*. Tradução do original latino por C. A. R. do Nascimento. 2ª ed. São Paulo: Editora Unesp, 2005.

SANTOS, C. A. dos. *O Górgias retórico e o* Górgias *de Platão*. Dissertação (Mestrado em Filosofia). São Paulo: Pontifícia Universidade Católica de São Paulo, 2008.

SPECA, A. *Hypothetical syllogistic and Stoic logic*. Leiden, Boston, Köln: Brill, 2001. (Philosophia Antiqua, 87)

A CONTINUIDADE ENTRE OS MÉTODOS FORMAIS E INFORMAIS DE DEMONSTRAÇÃO NA PRÁTICA FILOSÓFICA

Nastassja Pugliese[1]

RESUMO

O papel da lógica na filosofia é fundamentalmente metodológico já que é fazendo uso das diferentes estruturas de demonstração que teses filosóficas são estabelecidas. Dada a variedade dos modos de apresentação dos textos filosóficos, entendo que para que a tese do papel da lógica na filosofia - em sentido amplo - possa ser defendida, a lógica deve ser tomada sua concepção mais larga, abarcando tanto a lógica formal quanto a informal. Argumento que, mesmo considerando as várias formas de demonstração filosóficas e as diversas relações possíveis entre lógica e filosofia, há uma continuidade entre os métodos formais e informais de demonstração. Para mostrar esta continuidade, partirei da noção fundamental de inferência lógica para estabelecer os critérios de diferenciação entre o formal e o informal, e reconstruirei brevemente a história filosófica da lógica para mostrar suas transformações e seu distanciamento progressivo do lugar metodológico na filosofia. Finalmente, considerando a literatura recente sobre lógica formal e informal, aponto os problemas relativos às suas diferenças e, partindo do reconhecimento de que lógica cumpre um papel metodológico na filosofia, argumento que a lógica formal e informal são práticas que estão em continuidade e se beneficiam mutuamente. No que diz respeito ao modo como a lógica informal se beneficia da lógica formal, o artigo investigará as vantagens e limitações em se tomar os sistemas de dedução natural como instrumento de análise de demonstrações em linguagem natural. No que diz respeito aos modos como a lógica formal pode se beneficiar da lógica informal, procuraremos mostrar que avanços na lógica formal decorrem, além dos esforços de formalização de provas informais, de uma análise informal das estratégias de cálculo e dos princípios segundos os quais uma certa linguagem lógica opera.

Palavras-chave: Lógica Formal. Lógica Informal. Método Filosófico. História da Lógica. Inferência. Demonstração.

ABSTRACT

The role of logic in philosophy is fundamentally methodological since it is through the usage of different demonstrative structures that philosophical theses are established. Given the variety of modes of presentation of philo-

[1] Professora Adjunta de Filosofia da Educação na Faculdade de Educação da UFRJ e docente permanente do Programa de Pós-Graduação em Lógica e Metafísica do Departamento de Filosofia da UFRJ. E-mail: nastassja.saramago@fe.ufrj.br. ORCID: 0000-0002-1961-3873.

sophical texts, I understand that for the thesis of the methodological role of logic in philosophy - in an encompassing sense - to be defended, logic has to be taken in its most general conception, involving both formal and informal logic. I argue that, even considering the various forms of philosophical demonstration and the various possible relationships between logic and philosophy, there is a continuity between formal and informal methods of demonstration. In order to show this continuity, I will start from the fundamental notion of logical inference to establish criteria for distinguishing between the formal and the informal, and I will briefly reconstruct the philosophical history of logic to show its transformations and its progressive distancing of its methodological role in philosophy. Finally, considering the recent literature on formal and informal logic, I will point out problems with respect to their differences and, from the recognition that logic plays a methodological role in philosophy, I will argue that formal and informal logic are practices that are in continuity and are mutually beneficial. With respect to the way in which informal logic benefits from formal logic, the article will investigate the advantages and the limitations in taking systems of natural deduction as instruments of analysis of demonstrations in natural language. With respect to how formal logic benefits from informal logic, I will show that advances in formal logic depend upon not only efforts for formalising informal proofs, but also from an informal analysis of calculi strategies and from the principles according to which a logical language operates.

Key-words: Formal Logic. Informal Logic. Philosophical Method. History of Logic. Inference. Demonstration.

Introdução

Em uma investigação sobre a relação entre a lógica formal e informal a partir do lugar da lógica como método na filosofia, é preciso mostrar em que medida o estudo das noções de inferência e validade, das diferentes estratégias de prova e de demonstração se relacionam com a filosofia. Esse passo é importante porque a teoria da prova é motivada pelo estudo das estruturas gerais das provas do modo como elas ocorrem na matemática, mas as questões tratadas em teoria da prova e a própria noção de prova não se limitam e não são circunscritas à sua expressão na matemática. A lógica formal e a teoria da prova[2], por nos oferecem uma teoria sobre noções lógicas, sobre demonstrações,

[2] Há diversos modos de estabelecer as relações que aqui estabeleço já que há vários métodos dedutivos disponíveis e, mais importante ainda, várias formas de se fazer filosofia. Neste artigo darei exemplos a partir de desenvolvimentos vindos da teoria da prova e os relacionarei com toda prática filosófica que se proponha demonstrativa. Este trabalho não se propõe discutir o que é ou não é filosófico, isto é, não discute o problema da demarcação. A intenção é mostrar que, nas práticas filosóficas demonstrativas, a lógica cumpre um papel metodológico e que lógica formal e informal estão em continuidade neste tipo

derivações e as propriedades dos sistemas lógicos, elas nos informam também sobre as estruturas gerais dos argumentos e dos sistemas filosóficos. Assim, minha hipótese, neste artigo, é a de que o estudo dos métodos de prova na lógica e sua comparação com as demonstrações informais traz compreensão sobre o modo pelo qual construímos argumentos filosóficos. A lógica formal nos permitirá compreender como as relações de inferência ocorrem no interior de um sistema dedutivo e também, por oposição, como elas se dão fora dele. Assim, compreendendo como a lógica formal sistematiza procedimentos de raciocínio empregados em linguagem natural podemos a partir daí investigar se estas sistematizações se afastam ou mantêm proximidade com os processos de pensamento que ela, de alguma modo, codifica[3]. Este artigo, ao se debruçar sobre as noções de método, prova e demonstração, pretende comparar os procedimentos formais e os não-formais a partir do contexto da prática filosófica. Dado que não é evidente considerar que a lógica formal possa servir de método para a filosofia e esteja, portanto, em continuidade com procedimentos de raciocínio informais realizados em linguagem natural, mostrarei, como a noção de inferência é essencial para todas as noções de demonstração, sejam elas formais ou informais.

A noção de inferência é central para a demonstração filosófica, pois ela se encontra presente em todo e qualquer processo de raciocínio, seja ele um procedimento lógico formal ou informal. Há diferentes formas de inferência, mas para nossos fins precisamos apenas diferenciar entre consequência lógica (que é uma inferência num contexto dedutivo) e as inferências em geral (que ocorrem em contextos dedutivos ou não-dedutivos). A partir da idéia de que há inferências dedutivas e não-dedutivas, podemos também esclarecer o que entendemos como prova e como demonstração. De modo análogo, partimos do princípio de que a prova é diferente da demonstração na medida em que a prova é oferecida e realizada dentro de um contexto de dedução e as de-

---

de prática.

[3] Como mostra Shapiro (1998), "um slogan comum é o de que a lógica codifica normas pré-teóricas do raciocínio correto, como a noção de consequência lógica. Este slogan é certamente correto, dentro de suas limitações, mas ele também é certamente limitado. O que é esta noção pré-teórica de consequência lógica e o que quer dizer codificar alguma coisa?" (Shapiro 1998, p. 1). Em uma pesquisa futura procurarei investigar estas questões, mas aqui me limitarei ao problema da relação entre a lógica formal e a lógica informal.

monstrações podem vir a ocorrer também em contextos não-dedutivos. Assim, do mesmo modo, a noção de inferência é relacionada com a noção de consequência lógica, a demonstração é um conceito mais geral que inclui o conceito de prova, mas não se reduz a ele. Finalmente, consideraremos como método, certos procedimentos ordenados ou sistematizados para a realização de provas ou demonstrações a partir da análise, construção ou cálculo de inferências. Nossa investigação parte, portanto, dessas distinções fundamentais.

Inferência como conceito lógico-filosófico fundamental

Comecemos, portanto, analisando como a noção de consequência lógica, a versão forte da inferência, serve de fio condutor para o estudo da continuidade entre lógica formal e informal[4]. Griffiths (2014), no artigo "Formal and Informal Consequence", mostra que a definição padrão de consequência lógica é originária da teoria de modelos, mas chama atenção para o fato de que há, ao menos duas noções formais de consequência lógica: uma vinda da teoria de modelos e outra da teoria da prova[5]. Elas são as seguintes:

> (*def. teoria de modelos*): uma sentença ϕ é consequência lógica de um conjunto de sentenças Γ *sse* todo modelo de Γ é um modelo de ϕ.

[4] Shapiro (1998) aponta que há várias questões a serem investigadas a respeito da noção de consequência lógica no contexto da teoria de modelos. Dentre elas estão: questões sobre as relações entre as linguagens formais e as linguagens naturais, questões sobre a natureza da forma lógica, questões sobre as diferenças entre os termos na lógica e estes mesmos termos utilizados de forma não-lógica e, ainda, questões epistemológicas e metafísicas. Neste trabalho, trataremos da noção de inferência e como ela serve de fio condutor para analisar as semelhanças entre lógica formal e informal. Enquanto na lógica formal a inferência torna-se consequência lógica, na lógica informal, temos uma noção de inferência mais fraca, que é a noção informal de consequência.

[5] Tomamos este artigo de Griffths (2014) como base porque ele oferece definições de consequência lógica que colabora para nossos fins. Importante deixar claro que a literatura em filosofia da consequência lógica é extensa e inclui desenvolvimentos complexos tanto em teoria do significado (ex. Dummet 1991) na construção de sistemas lógicos (Shoesmith & Smiley 1978, Brandon 1994). Os autores mais tradicionais para as discussões sobre o tema no contexto da lógica moderna são Bolzano e seu conceito de dedutibilidade lógica (1837) e Tarksi (1986). Sabemos, no entanto, que a noção de consequência lógica é preocupação comum entre todos os filósofos que contribuíram direta ou indiretamente para a história da lógica. Neste artigo, além destas duas noções de consequência lógica, ainda consideramos que há uma versão fraca da noção de consequência, instrumentalizada na lógica informal, e que captura nossa noção pré-teórica de inferência não-dedutiva. É a partir desta noção de informal de inferência que mostraremos a continuidade entre lógica formal e lógica informal.

> (*def. teoria da prova*): uma sentença ϕ é consequência lógica de um conjunto de sentenças Γ *sse* ϕ for dedutível de Γ usando as regra de um sistema de prova padrão para a lógica de primeira ordem.

A primeira definição formal de consequência lógica é semântica enquanto a segunda é sintática, mas ambas capturam o seguinte resultado: se as sentenças de Γ forem tomadas como as premissas de um argumento e ϕ como a conclusão deste argumento, temos que, se ϕ for consequência lógica de Γ, então podemos dizer que o argumento é dedutivamente válido e sólido. Assim, tomando conjuntamente estas definições de consequência lógica, podemos compreender porque o "bom argumento" é considerado como aquele que é válido e sólido. O "bom argumento" é válido e sólido porque, dada sua forma, nele a conclusão segue das premissas de modo que ela é consequência lógica das premissas, ou seja, é impossível que, se as premissas forem verdadeiras, a conclusão seja falsa. Dizer que Γ tem modelo, é dizer que há ao menos uma interpretação possível na qual todas as sentenças de Γ são verdadeiras. Assim, dizer que ϕ é consequência lógica de Γ *sse* todo modelo de Γ é um modelo de ϕ, quer dizer que se todas as sentenças de Γ forem verdadeiras em determinado modelo, então ϕ também será verdadeiro neste modelo. Dizer, por outro lado, que de Γ deduz-se ϕ a partir de um sistema de prova padrão, é dizer que, neste sistema, ϕ é derivável do conjunto de sentenças Γ. Assim, tomando Γ e ϕ como partes de um argumento, dizer que ϕ é derivável do conjunto de sentenças Γ, é o mesmo que dizer que o argumento é válido; e dizer que há um modelo no qual tanto as sentenças de Γ quanto ϕ são verdadeiros ao mesmo tempo, é dizer que o argumento é sólido. Deste modo, vemos que as definições formais de consequência lógica vindas da teoria de modelos e da teoria da prova explicam certas propriedades dos argumentos.

Argumentos, entretanto, não se restringem a este formato onde a conclusão é consequência lógica das premissas. Argumentos, de um modo mais geral, são conjuntos de sentenças nos quais algumas são premissas e uma outra é a conclusão. Um argumento bem formado é aquele cuja conclusão se segue ou provavelmente se segue das premissas. Quando dizemos que a

conclusão segue das premissas, dizemos que é possível inferir a verdade da conclusão a partir da assunção de que as premissas são verdadeiras. Quando a verdade da conclusão necessariamente segue da verdade das premissas, temos um argumento dedutivo onde a conclusão é consequência lógica das premissas. Nos casos em que a verdade da conclusão de um argumento é apenas provável dada a verdade das premissas, não podemos mais dizer que a conclusão é consequência lógica das premissas. No entanto, dado um argumento bem formado, há, necessariamente, uma relação de implicação entre as premissas e a conclusão. Essa relação de implicação é chamada, como apontamos acima, de relação de inferência.

Assim, de um modo mais abrangente, podemos retomar a definição de argumento e recolocá-lo do seguinte modo: um argumento é um conjunto de sentenças no qual algumas destas sentenças são premissas, uma delas é a conclusão, de modo que há uma relação de inferência entre as premissas e a conclusão. Ao se incluir, de modo explícito, a noção de inferência na definição de argumento, torna-se possível categorizar argumentos de acordo com os diferentes graus de necessidade com os quais a verdade da conclusão se segue as verdades das premissas. Como vimos nas definições, a consequência lógica é uma versão forte da inferência. Em um argumento dedutivo cuja conclusão é consequência lógica das premissas, a conclusão pode ser deduzida da verdade das premissas a partir de um procedimento de prova (no caso, demonstrado através de uma derivação). Isto porque, neste caso, a verdade da conclusão segue necessariamente da verdade das premissas. A relação de inferência mais fraca é aquela na qual a conclusão é provavelmente verdadeira e há um certo risco em concluir a verdade da conclusão segue-se da verdade das premissas. Assim, temos que a relação de inferência é mais abrangente que a relação de consequência lógica, já que a consequência lógica é um tipo de relação de inferência mas nem toda relação de inferência é uma consequência lógica. E, neste último caso, no caso dos argumentos não-dedutivos, ainda que possamos demonstrar que a verdade da conclusão provavelmente se segue da verdade das premissas, não é possível fazer uso de procedimentos de prova para realizar esta demonstração. Esta demonstração deve ser feita de modo informal, a

partir da enumeração de razões que justifiquem a cogência de tal argumento, por apelo às noções comuns ou às ferramentas da teoria de probabilidades.

Estas distinções nos servem para que possamos observar como há uma diferença de grau entre os tipos de inferência e que é correto dizer que, assim como a consequência lógica é a forma mais forte de inferência, o argumento dedutivamente válido é a forma mais forte de um argumento. Outra consequência destas diferenças entre os tipos de relação de inferências que definem as formas dos argumentos é que deve haver um método de demonstração distinto para cada forma de argumento. O método para demonstrar que a conclusão de um argumento dedutivamente válido segue-se da verdade de suas premissas não é e não pode ser o mesmo que o método utilizado para demonstrar que a conclusão de um argumento não-dedutivo provavelmente segue-se da verdade das suas premissas. Assim, para uma compreensão e um estudo mais detalhado das dimensões metodológicas da lógica na filosofia, é preciso analisar as diferenças entre as diferentes propriedades dos argumentos, os diferentes métodos de demonstração e os vários graus de certeza das inferências. Vejamos como essas noções e estas diferenças foram operadas ao longo da história e marcaram o movimento de abstração crescente em lógica.

Breves considerações históricas e a lógica como método

A conclusão a qual chegamos, de que há diferentes graus de certeza nas relações de inferência, onde a consequência lógica é o tipo de inferência mais forte (dado que é dedutiva) está de acordo com (e de certo modo repete) o que Aristóteles argumenta nos *Segundos Analíticos*. Nele, Aristóteles defende que há apenas um tipo de demonstração, entre várias, que deve contar como uma demonstração verdadeiramente científica. A demonstração que, para Aristóteles, funciona como uma prova, ocorre apenas quando apresentamos um argumento na forma do mais forte dentre todos os silogismos possíveis. Aristóteles argumenta que há, entre os silogismos válidos, o mais seguro deles. Este "silogismo entre os silogismos" é chamado, posteriormente, de de-

monstração *potissima*[6] e só ela é considerada como uma demonstração verdadeiramente científica. Uma demonstração *potissima* é um argumento válido que tem a forma de um silogismo de primeira figura com premissas universais, mais conhecidas e anteriores que a conclusão. A conclusão desde silogismo, por sua vez, também é universal e está relacionada com as premissas do mesmo modo com que efeitos se relacionam com causas. Aristóteles estabelece que uma demonstração, para ser científica, deve observar certos requerimentos precisos. Importante deixar claro que estes requerimentos exigem muito mais do que a presença de uma relação de consequência lógica entre as premissas e conclusão de um argumento: a verdade das premissas de uma demonstração *potissima* deve ser imediata, primeira e anterior, ou seja, devemos conhecer a verdade das premissas antes e melhor do que a conclusão, e as premissas devem ser sobre as causas que explicam um certo efeito sobre o qual versa a conclusão. Já que para Aristóteles, conhecer é conhecer pelas causas (*aitia)*, o conhecimento científico deve ser obtido pela demonstração da relação entre causa e efeito. Com estes requerimentos fortes, Aristóteles dá os primeiros passos na direção do uso da lógica como método de demonstração científica (ou como método de prova)[7]. Aristóteles, apontando as diferenças entre o raciocínio científico e o raciocínio retórico, e entre os diversos tipos de silogismos, nos indica que para fazermos ciência (e filosofia) precisamos analisar as diferentes propriedades dos argumentos. Estas caracterizações dadas por Aristóteles nos permitem compreender como a defi-

[6] A silogística como teoria da investigação e da demonstração manteve-se, durante séculos, como o método mais seguro para as práticas científicas e matemáticas. Na escolástica, comentadores de Aristóteles argumentavam que as ciências mais exatas, como a matemática, eram exatas porque suas demonstrações estavam em conformidade com a definição de ciência perfeita de Aristóteles. Ou seja, tinham suas verdades demonstradas por silogismos demonstrativos. Este tipo de demonstração entre as demonstrações, ou este silogismo entre os silogismos, foi chamado por Piccolomini, no século XVI, de demonstração *potissimae (Barbara)*. Piccolomini reconhece que a teoria de Aristóteles, de todo modo, ofereceu as categorias conceituas que permitiram a análise do discurso científico, matemático e lógico. Entretanto, na transição do século XVI para o século XVII, a concepção de ciência de Aristóteles é questionada em um movimento chamado *quaestio certitudine mathematicarum* (Mancosu 1996). As críticas à Aristóteles e a constatação da impossibilidade de se realizar demonstrações *potissimae* em todas as áreas da matemática e do saber, levaram à construção de novos métodos de investigação, justificação e demonstração. Nesta esteira de desenvolvimento, se insere o esforço de Leibniz, por exemplo, pela algebrização das demonstrações. A lógica moderna, portanto, pode ser compreendida por suas semelhanças e diferenças em relação à Aristóteles e sua teoria da investigação e da demonstração.

[7] A motivação por trás do estabelecimento do sistema lógico de Aristóteles e sua teoria do silogismo é compreender e realizar demonstrações (*apodeixis*), ou seja, "deduções que produzem conhecimento". A noção de demonstração em Aristóteles e a teoria dos silogismos funcionam como prelúdio metodológico de sua filosofia.

nição de propriedades e noções lógicas nos permitem codificar noções pré-teóricas e sistematizar métodos de demonstração. De um modo mais geral, as distinções de Aristóteles também nos auxiliam a compreender os desenvolvimentos metodológicos na história da filosofia e sua relação com os avanços da lógica. Com Aristóteles, vemos que a busca pelas melhores estratégias de demonstração é um esforço que se confunde, em um certo sentido, com o começo da filosofia[8].

É importante salientar que Aristóteles concebe não só as propriedades e noções lógicas fundamentais, mas também a primeira sistematização de procedimentos de demonstração ao categorizar as formas válidos dos silogismos. A lógica e a teoria do silogismo de Aristóteles foram tão importantes que, por séculos, tornaram-se a grande referência para as práticas científicas e matemáticas. A força do sistema lógico aristotélico é tanta que Kant, ao descrever um momento na história anterior às transformações da lógica matemática no século XIX, concluiu que Aristóteles descobriu tudo o que havia para ser descoberto em lógica e que não havia mais desenvolvimento possível para além da lógica aristotélica. Kant realiza essa leitura mesmo depois da ocorrência das querelas matemáticas do final do século XVI, início do século XVII e os avanços na lógica matemática por parte de Leibniz. Na tradição escolástica, comentadores de Aristóteles argumentavam que as matemáticas eram tidas como ciências exatas porque suas demonstrações estavam em conformidade com a definição de ciência perfeita de Aristóteles. No século XVII, as obras da matemática clássica começaram a ter seus resultados, antes considerados absolutamente certos, questionados. Piccolomini, por exemplo, procura mostrar que as demonstrações matemáticas não podem ser demonstrações *potissima* porque as relações matemáticas não podem ser pensadas em termos causais. De todo modo, como afirma Mancosu, "a concepção aristotélica de ciência ofereceu as categorias conceituais que permitiram a análise do discurso matemático no começo do século XVII" (Mancosu 1996

[8] Como nos mostra Kneale & Kneale (2008, p. 23), "apesar da palavra 'lógica' ter encontrado seu significado moderno apenas 500 anos depois quando foi utilizada por Alexandre de Afrodísias, o escopo do estudo depois chamado de lógica foi determinado pelo conteúdo do *Órganon*" [de Aristóteles]. Como vimos, Aristóteles desenvolveu sua lógica com o objetivo apresentá-la como um órganon do conhecimento, ou seja, como um método seguro a partir do qual podemos construir demonstrações científicas.

p.10). O rápido desenvolvimento e o esforço de algebrização das matemáticas mudaram tanto a cara da matemática como da lógica e da filosofia. No século XVII, a influência do método matemático, inspirado na sistematização iniciada com a lógica aristotélica, se fez valer em outras disciplinas como a filosofia, a teologia e a filosofia natural.

O século XVII é a época onde se vê de modo mais evidente a influência do método lógico-matemático na filosofia. Com isto, não queremos dizer apenas que houve uma multiplicação das aplicações de técnicas lógico-matemáticas em certas áreas de estudo específicas (como fez Galileu na matematização da mecânica e Descartes na algebrização da geometria; cf. Mancosu 1996), mas que a matemática é tomada como exemplo de ciência segura. Entendida como um exercício de clareza, ordem, simplicidade, o método lógico-matemático era utilizado para dar segurança e certeza às investigações de diversas áreas do conhecimento. Na filosofia, o caso mais emblemático é o de Espinosa ao escrever uma ética (a partir de seu sistema metafísico) demonstrada à maneira dos geômetras. Neste esforço pela matematização das ciências, o lugar da lógica aristotélica começa a ser questionado até que, com os desenvolvimentos da lógica moderna no século XIX, a teoria dos silogismos é praticamente abandonada[9]. Entretanto, Smith (2017) nos lembra que

> durante o avanço da lógica formal moderna com Frege e Pierce, aqueles que aderiam à lógica tradicional (que viam a si mesmos como descendentes da lógica aristotélica) e aqueles que aderiam à nova lógica matemática viam-se como rivais que defendiam noções incompatíveis de lógica. Pesquisas mais recentes, entretanto, aplicam várias técnicas da lógica matemática às teorias de Aristóteles revelando (na opinião de muitos) um grande número de semelhança de abordagem e de interesse entre Aristóteles e os lógicos modernos. (Smith 2017)

Entretanto, do mesmo modo que hoje podemos ver as semelhanças de abordagem e interesse entre a lógica aristotélica e a moderna a partir da perspectiva da lógica atual, podemos também investigar os pressupostos da lógica moderna a partir de sua dimensão metodológica ou propedêutica na filosofia.

[9] Como aponta Mancosu (1996), há duas vertentes explicativas da relação entre a lógica aristotélica e os desenvolvimentos que se iniciam no século XVII e culminam no XIX com a lógica moderna: a perspectiva da continuidade (Crombie 1959, Randal 1961, Von Wright 1971) e a da descontinuidade (McMullin 1965, Gilbert 1963). Em algum momento futuro, a presente pesquisa se debruçará sobre este tema. Mas esta não é a nossa preocupação presente.

Ou seja, acreditamos que podemos (e devemos) analisar a lógica moderna como sendo (ainda) parte da dimensão metodológica da filosofia. Isto porque a lógica moderna nos informa sobre estruturas de argumentos e demonstrações utilizadas no processo de conhecimento ainda que ela tenha se voltado a estratégias de cálculo que independem do processo de conhecimento filosófico.

Esta busca por compreender o lugar da lógica moderna na metodologia filosófica se confunde com as motivações de Aristóteles para construir sua lógica porque quando ele concebe uma ciência dedutiva a partir de certas definições iniciais, sua meta fundamental "não é um guia prático para a argumentação, mas o estudo das propriedades dos sistemas de inferência" (Smith 2017). Este estudo meta-teórico sobre a lógica permite à Aristóteles explorar as relações entre demonstração e conhecimento, entre inferência e verdade, fazendo uso da lógica como método próprio para a ciência e como instrumento da filosofia. Com o surgimento da lógica moderna, voltada para a matemática, perde-se a evidência da relação fundamental que existe entre lógica e as diversas ciências. Neste artigo tomamos a lógica a partir destas relações e nosso interesse é, portanto, tanto no aspecto prático quanto no aspecto histórico da lógica. A história da lógica permite que tenhamos a visão de conjunto necessária para a análise das práticas atuais da lógica formal e para que possamos investigar suas possibilidades metodológicas no ensino de filosofia. Deste modo, é importante lembrar os marcos fundadores da lógica moderna (lógica simbólica ou lógica matemática) e as transformações que ela sofreu nos últimos séculos.

Acreditamos que a lógica moderna cumpre uma função metodológica não apenas por oferecer uma teoria dos argumentos válidos com mais poder expressivo que a de Aristóteles, mas também por nos fornecer uma compreensão sobre o cálculo como operação fundamental do pensamento (Grattan-Guinness 2004). A lógica formal moderna, também chamada de lógica simbólica ou lógica matemática, surge no século XIX como um desenvolvimento a partir da matemática tradicional e de sistemas lógicos anteriores[10]. Sabemos

[10] A lógica aristotélica foi abandonada em suas pressuposições metafísicas e epistemológicas, mas não em sua estrutura. As formas de inferência válidas na lógica aristotélica, aparecem na lógica formal moderna como casos especiais, ou seja, com um subconjunto de formais mais gerais do raciocínio.

que a lógica formal, ou a lógica simbólica, realiza importantes avanços frente à lógica aristotélica na medida em que ela encontra um modo de tratar tanto de quantidades quanto de estruturas relacionais de alto grau de complexidade. O desenvolvimento da lógica formal se deve, principalmente, ao trabalho de Boole e de Frege. Boole, no *The Mathematical Analysis of Logic* (1847), construiu um sistema de cálculo inspirado em técnicas da álgebra capaz de formalizar as regras da silogística tradicional. Esta monografia foi fundamental porque nela Boole foi bem sucedido em mostrar que as as operações da lógica aristotélica são análogas às operações algébricas. Esta semelhança entre o raciocínio o lógico e o cálculo aritmético já tinham sido exploradas por Leibniz, que tinha a intenção de construir um método mecânico para a análise de argumentos. Leibniz procurou explorar as analogias existentes entre a operação algébrica da adição e a operação lógica da conjunção (Esquisabel 2012, Casanave 2012). Mas, ao fazê-lo, Leibniz concluiu que esta forma de compreender a lógica como fundada em operações análogas às da álgebra encontra limitações: quando representamos, por exemplo, a conjunção pelo sinal de adição, chegamos à falsidades algébricas (como A + A = A) e limitamos as possibilidades de combinações das operações lógicas. A estratégia de Boole possui limitações semelhantes, mas, ainda assim, ele realizou um avanço significativo na lógica formal ao construir um algoritmo em linguagem algébrica possível de ser aplicado a uma variedade de argumentos de diferentes níveis de complexidade. Até então, o método tradicional, baseado nas estratégias de Aristóteles, se resumia à catalogação das forma de silogismo válido em suas formas mais simples. Boole desenvolveu um cálculo de operadores de classes capaz de representar algebricamente as formas dos silogismos válidos, tomando os símbolos da álgebra como representações de operações lógicas[11]. Boole então mostrou as vantagens de se investigar as semelhanças entre a lógica e matemática, trazendo novamente a questão filosófica sobre a relação entre a lógica e ontologia. Isto porque, se a lógica e a matemática compartilham de operações fundamentais, então é possível que as leis lógicas sejam também as

[11] Como mostra Ewald (2007), o sistema de Boole, entretanto, não permitia operar com relações para além daquelas existentes entre sujeito e predicado em uma proposição e também não possuía quantificadores.

leis do pensamento. Assim, os avanços algébricos na lógica moderna retoma questões fundamentalmente filosóficas como, por exemplo, o estudo sobre os comprometimentos ontológicos (ou ausência deles) subjacentes aos sistemas lógicos.

O segundo grande avanço que constitui a virada matemática na lógica foi realizado por Frege (Grattan-Guiness 2004). Frege construiu o primeiro sistema formal capaz de realizar cálculos de predicados em sentenças quantificadas. Zalta (2017) caracteriza o avanço realizado por Frege como "o desenvolvimento de um método mais perspícuo de representar formalmente a lógica de pensamentos e inferências" fornecendo, deste modo, os fundamentos da lógica moderna. Frege não só construiu um sistema formal capaz de analisar sentenças complexas e quantificadas, mas ofereceu também, como fundamentação teórica, uma teoria da prova e da definição. O cálculo de predicados de Frege é um sistema formal, incluindo uma linguagem formal e um método de prova, no qual podemos representar inferências válidas que ocorrem entre sentenças nas quais propriedades são predicadas de objetos (Zalta 2017). A formulação de Frege é muito próxima do cálculo de predicados moderno[12]. Se tomamos uma sentença e substituímos um nome completo que aparece na sentença por um marcador de lugar, então temos como resultado uma expressão incompleta ou um conceito. Um conceito para Frege é um certo tipo de função que mapeia qualquer argumento em um valor de verdade (verdadeiro ou falso). Por exemplo, a sentença '4 é par' deve ser formalmente representada como P(4), onde o predicado 'é par' é analisado como denotando o conceito P( ) tal que P ( ) relaciona todos os números pares com o verdadeiro e todos os outros como falso. A predicação '4 é par' é analisada por Frege como se ela fosse uma função onde 'é par' é o conceito e '4' é o argumento. Assim, o sistema de Frege permitiu a análise de predicados não-matemáticos como "é vermelho" ou "é verde". Frege considera que sua lógica fornece todos os princípios para realização de provas dedutivas. Como expresso na *Begriffsshrift*: "Tudo o que é necessário para a realização de uma inferência correta é completamente expressado, mas o que não é necessário geralmente não está

[12] Na lógica de Frege todos as expressões completas são termos que denotam e, por isso, sua lógica é chamada também de lógica de termos.

indicado; nada é deixado para ser adivinhado" (Frege 1879 p.3)[13]. Assim, enquanto Aristóteles ofereceu um método de combinação de formas válidas previamente categorizadas, Frege concebeu um sistema fechado e finito de regras capaz de funcionar como um algoritmo para a análise de uma quantidade infinita de argumentos válidos com diferentes níveis de complexidade. Dos Santos resume a caracterização de lógica feita no *Nachgelassene Schriften* (1969 pp. 3 e p. 190) por Frege do seguinte modo: "A tarefa específica da lógica pode ser caracterizada (…) como a de estabelecer sistematicamente leis aptas a medir a legitimidade de conclusões (*Schlüsse*) ou demonstrações (*Beweise*)"[14]. O sistema de Frege permitiu a manipulação de um conjunto de fórmulas consistentes em uma dada linguagem de modo que, sob um ponto de vista meramente combinatório, fossem produzidas derivações de uma dada fórmula a partir de outras fórmulas assumidas como axiomas ou mesmo como meras hipóteses[15].

Sobre a lógica informal e sua relação com a lógica formal

A motivação por trás desta breve reconstrução histórica é mostrar que o desenvolvimento da lógica pode ser lido como um esforço, cada vez mais complexo, de formalização de procedimentos de inferência que realizamos quando raciocinamos[16]. Este fato nos permite observar a relação frutífera existente entre demonstrações formalizadas e não-formalizadas e a necessidade de se estudá-las a partir de uma perspectiva de continuidade entre estes esforços.

---

[13] A lógica de Frege também inclui a análise de conceitos. Neste artigo não nos debruçaremos sobre esse aspecto da lógica fregeana (como também não nos debruçamos sobre a teoria das definições de Aristóteles), apesar de reconhecermos que este aspecto da lógica, enquanto análise conceitual, é de suma importância para a filosofia.

[14] Dos Santos, Luiz Henrique Lopes. *O Olho e o Microscópio.* Nau Editora. Rio de Janeiro, 2008. p.23

[15] Esta realização inspirou o artigo de Hilbert de formalização da matemática a partir da escolha inicial de certos axiomas. Este artigo de fundamentação da matemática na lógica motivou diversos avanços, mas a demonstração, por Gödel, de sua impossibilidade gerou a crise que está na base do desenvolvimento da teoria da prova do começo do século XX.

[16] Importante salientar as limitações desta reconstrução tanto no que diz respeito às caracterizações das descontinuidades nos desenvolvimentos da lógica quanto no que concerne às contribuições feitas por mulheres e autores não-canônicos à história da lógica. Em pesquisas futuras, a serem publicadas (Pugliese & Secco *no prelo*), procurarei apontar os problemas com estes tipos de narrativas históricas nas quais, com este presente trabalho, eu mesma me incluo. Há investigações importantes (e demoradas) a serem feitas nesta direção, como por exemplo, a análise do modo com a história da lógica é apresentada em livros didáticos em Português. Começamos a realizar este tipo de investigação em Secco & Pugliese (2016), mas há muito o que ser investigado ainda.

A meta mais ampla desta investigação e que não se resolverá neste breve artigo é de investigar como a noção de prova, enquanto ferramenta de análise de sistemas axiomáticos, pode informar e auxiliar a análise de teorias filosóficas não-axiomáticas e não-axiomatizáveis. Ou seja, este é um estudo propedêutico para que posteriormente continuemos a avaliar a possibilidade de se tomar os princípios da prova, como a explicitação de regras de inferência e a escolha de axiomas, como instrumentos para a avaliação de argumentos de diversos tipos (dedutivos e não dedutivos). Para mostrar a interseção entre lógica formal e a prática de argumentação filosófica, nos deteremos agora sobre a lógica informal.

A lógica informal, considerada como em oposição à lógica formal, não parece ter uma prática ou metodologia bem definida. Isto porque, enquanto a lógica formal realiza estratégias mecânicas para a análise de argumentos dedutivamente válidos, a lógica informal cumpre um papel de sistematização e análise, mas de um modo não-mecânico. Entretanto, só encontramos dificuldade em definir a lógica informal se a pensarmos como um substituto da lógica formal. Considerada a partir desta oposição, a lógica informal não passa de uma lógica formal mal feita[17], utilizando os princípios da dedução da lógica tradicional para operar sobre argumentos não-dedutivos. Segundo estes princípios, os argumentos dedutivamente válidos são os únicos bons argumentos e a inferência mais genuína é a realizada por dedução. Nesta perspectiva dedutivista da lógica informal (Groarke 1999), privilegia-se a estratégia de tradução-cálculo para a manipulação e análise de argumentos, na qual todo e qualquer argumento deve ser transformado em um outro, semelhante ao original, mas que tenha a forma de um argumento dedutivamente válido para que então se realize a análise de inferências. Entretanto, este privilégio da estratégica de tradução-cálculo limita as potencialidades metodológicas da lógica informal e suas possíveis formas de interação tanto com a lógica formal quanto com a linguagem natural para fins de análise de argumentos e modos de inferência.

[17] Como diz Fogelin (1999), a lógica informal não passa de uma "informal formal logic" quando ela é concebida a partir do que ele chama de "preconceitos dedutivistas". Aqui não trataremos de preconceitos dedutivistas, mas de uma visão dedutivista que desafia o bom funcionamento da lógica informal.

Para mostrar que a perspectiva da oposição ente lógica formal e informal não é a forma mais fértil de abordar estas práticas e suas relações com a filosofia, apontarei o problema da definição da lógica informal (Hintikka 1989) para então descrever o método informal de análise lógica de argumentos mais adotado (Feldman 1999, Hughes et al 2015, Bonevac 1999, Morrow & Weston 2011). A partir deste método de análise de argumentos na lógica informal, bem como os princípios que guiam a análise e a busca por padrões de argumentos não-dedutivos que estruturam esta prática, argumentarei que a lógica informal cumpre um papel distinto do da lógica formal. Assim, a partir do estudo desta prática geral de análise de argumentos, mostrarei que a lógica informal não é oposta à lógica formal, mas está em necessária continuidade com ela. Esta compreensão da lógica informal permite a abordagem complementar (Johnson 1999) de ambas as disciplinas de modo que uma se beneficie das práticas da outra, sem que haja conflito de metodologias e confusão de escopo.

As controvérsias entorno da chamada "lógica informal" se justificam porque a lógica informal, por não oferecer um modelo para dar conta de demonstrações e inferências através de regras e procedimentos dedutivos, não é exatamente uma lógica[18]. Ao mesmo tempo, a lógica informal constrói e categoriza procedimentos alternativos de demonstração e inferência na tentativa de sistematizá-los, não sendo, portanto, exatamente um esforço "informal"[19]. Ou seja, as dificuldades entorno da lógica informal existem porque ela não é nem exatamente uma lógica e nem um procedimento não sistematizado ("informal"), sendo um termo genérico o bastante para causar incômodo nos mais diferentes contextos. Hintikka, por exemplo, no artigo "The Role of Logic in Argumentation", de 1989, parte do princípio de que expressão "informal logic" é um solecismo, uma cacologia, um oxímoro e defende a tese forte de que não há nenhuma teoria geral, respeitável, sobre a argumentação infor-

[18] O conceito de lógica é central aqui dado que o argumento dependente de uma certa concepção de lógica. Ainda que seja possível e, quiçá, inclusive necessário, definir a lógica de modo mais amplo, no presente contexto, por lógica entendo a lógica formal como cálculo proposicional e o cálculo de predicados de primeira e segunda ordem. Ou seja, como a tradicional lógica clássica.

[19] Novamente, aqui entendo informal em um sentido estritamente não-técnico, ou seja, enquanto uma prática não-regimentada e sem regras de funcionamento explícitas.

mal[20]. Hintikka argumenta pela impossibilidade da lógica informal, pois exige que, para ser uma empreitada bem-sucedida, a lógica informal deve oferecer uma teoria da inferência que não seja limitada à inferência dedutiva mas que a inclua, de modo que a lógica informal possa, então, servir de substituto para a lógica formal. Ou seja, ou a lógica informal oferece uma teoria da inferência e uma metodologia de análise da mesma mais forte do que a teoria da inferência dedutiva e os procedimentos da lógica formal, ou ela não vale nada[21].

Tal expectativa é compreensível na medida em que defender o emprego de qualquer procedimento alternativo às estratégias de tradução e cálculo só se justifica se tais procedimentos tiverem um alcance igual ou maior que, por exemplo, os da lógica tradicional. As fraquezas da lógica informal, segundo Hintikka, são três. Nesta proposta de uma "teoria da inferência mais substancial": (1) não estão bem-definidas quais tipos evidências nos levam a realizar inferências não-dedutivas e quais as outras estratégias empregamos face a estas evidências; (2) não se apresentam explicações sobre os elementos que fazem da concepção tradicional de lógica uma boa teoria geral do raciocínio; (3) e, apela-se para critérios como escolher "a melhor explicação entre as possíveis", o que é um critério vago e fraco para sustentar qualquer teoria genuína. Entretanto, a não ser a segunda, estas alegadas fraquezas fazem parte da natureza dos argumentos tratados pela lógica informal e são os desafios mesmos que qualquer teoria sobre argumentos não-dedutivos deve propor-se enfrentar.

Gostaria de mostrar que o modo mais interessante de se abordar a lógica informal é encará-lo não como um substituto da lógica formal (o que se-

[20] "Unfortunately, there exists by any reasonable standard no respectable general theory of informal argumentation. It serves no constructive purpose to argue in detail for this judgment here. If your intellectual tastes are such that you are satisfied with the theoretical level of traditional rhetorical theories or of so-called theories of informal reasoning, I am not going to be able to re-educate you in half-an-hour." p. 4

[21] Segundo sua caracterização, a lógica informal pretende oferecer uma teoria do raciocínio de modo a não reduzi-lo a uma cadeira de inferências dedutivas. A lógica informal, segundo Hintikka, deveria oferecer uma teoria da inferência mais substancial, capaz de elencar e explicitar o conjunto de princípios do raciocínio que nos levam a mudar de crença quando somos apresentados a novas evidências (e não apenas o princípios que nos levam a inferir, de modo válido, que certas crenças se seguem necessariamente dada a verdade de outras). Até então, segundo ele (o artigo é de 1989), a lógica informal ainda não havia apresentado tal teoria da inferência alternativa à dedução. E, do modo como Hintikka interpreta o alcance desejado pela lógica informal, o "movimento" não apresentou uma alternativa desta envergadura até hoje e, penso eu, que nem irá apresentar. Hintikka espera que a lógica informal, para ser bem-sucedida, tenha um alcance igual ou maior que o da lógica formal.

ria impossível) e nem como um estudo sobre procedimentos de maior ou igual alcance que o dos sistemas dedutivos, mas como um esforço de sistematização de modos de argumentar e demonstrar que estão em continuidade com a dedução e que, de uma forma ou outra, a tomam como o modelo de inferência para o raciocínio correto. Assim, se por um lado, Hintikka exige critérios excessivamente rígidos para o bom funcionamento da lógica informal, por outro lado, ele está certo em esperar que ela torne claro seu lugar frente à lógica formal. O desafio posto por Hintikka é, portanto, mostrar que a lógica informal é uma empreitada viável e, ainda que não seja e não venha a ser um substituto da lógica formal, ela não deve ser identificada à falta de rigor e à inutilidade. O presente artigo, ao abordar as noções de método, prova e demonstração, pretende conseguir mostrar que a perspectiva da continuidade entre a lógica formal e a informal é a mais fértil para se compreender a dimensão metodológica da lógica na filosofia.

Sabemos que, se por um lado, não é necessário uma teoria exaustiva da inferência não-dedutiva para o bom funcionamento da lógica informal, por outro, é preciso que a lógica informal se posicione frente a lógica formal já que ela se define por contraste com ela. Além disso, dado que a lógica informal é uma disciplina em construção e tem como objeto argumentos realizados em linguagem natural, a lógica informal carrega consigo características mais próximas da filosofia do que da matemática. Deste modo, a discussão a cerca do melhor critério para se acessar um argumento será sempre dependente de julgamentos que envolvem a ponderação de razões, levando a critérios do tipo "a melhor explicação entre as possíveis". Juntamente com Johnson (2000) e Haack (1998), considero que a lógica informal e a lógica formal são contínuas e, portanto, complementares. A lógica formal é o estudo de argumentos dedutivamente válidos, ou seja, argumentos em que se as premissas forem verdadeiras, a conclusão se segue necessariamente.

Na lógica formal utilizamos sistemas dedutivos que nos permitem manipular argumentos dedutivamente válidos de alto grau de complexidade a partir de regras de inferência, derivação e transformação (sendo elas mesmas tautologias, teoremas dentro do sistema ou padrões de argumentos dedutiva-

mente válidos mais simples). A lógica informal, por sua vez, é um esforço de categorização e sistematização de argumentos em linguagem natural que são, em sua maioria, não-dedutivos. A lógica informal propõe um método geral de reconstrução e análise de inferências, mas este método não é um procedimento mecânico capaz de calcular o que se segue do quê, dado um conjunto qualquer de sentenças. Enquanto a lógica formal lida com as inferências que se seguem necessariamente a partir de um conjunto de sentenças supostamente verdadeiras, a lógica informal trata de inferências que se seguem apenas provavelmente a partir de um conjunto de sentenças supostamente verdadeiras. Dado que a lógica formal e informal ambas tratam do estudo e da análise de inferências, não há entre elas uma diferença de natureza, mas de grau de confiabilidade e de grau de certeza no que diz respeito a seus objetos e métodos. Como são ambas os estudos das inferências, lógica formal e lógica informal devem ser consideradas como uma unidade, ainda que cada uma cumpra funções distintas na análise de inferências. Essa unidade se dá porque ambas são análises do raciocínio, estudos do que segue do quê, ou seja, sistematizações de esquemas de inferências bem aceitos. Ainda que empreitadas contínuas, a lógica formal e a informal fazem usos de métodos distintos. Estes métodos, entretanto, são complementares. Para demonstrar sua complementaridade, é preciso mostrar duas coisas: que a lógica formal complementa a lógica informal e que a lógica informal complementa a lógica formal.

A lógica formal complementa a lógica informal porque as linguagens da lógica proposicional e de predicados são fundamentais para a compreensão do que é a relação de inferência correta, para o conhecimento do conceito de derivação e implicação, bem como para o compreensão do que são as regras do raciocínio e as formas lógicas propriamente ditas. A lógica formal é o núcleo duro do estudo das relações de inferência e é a partir dela que outros estudos são desenvolvidos. A lógica informal se beneficia da lógica formal na medida em que a lógica formal oferece os métodos mais fortes e os critérios mais bem-estabelecidos para a análise de inferências. A lógica formal oferece um modelo de funcionamento para a lógica informal, porque nela o método é bem formado quanto os objetos são bem estabelecidos, a ponto de, como vi-

mos, haver o contra-movimento que considera que apenas a lógica formal oferece uma verdadeira análise de inferências pois as únicas inferências que valem debruçarmo-nos sobre são as dedutivas. Como mostra Seoane (2014) a lógica formal pode nos auxiliar na compreensão da complexidade dos argumentos filosóficos por sua capacidade de expressar a riqueza das relações de inferências a partir do uso da estratégia de tradução e cálculo. Quando é possível utilizar esta estratégia, ela se mostra muito útil:

> dado um argumento em linguagem natural, o transformamos (via modelo diádico) em um conjunto de fórmulas de uma linguagem formal lógica e aplicamos a tal conjunto, na melhor das hipóteses, algum algoritmo que permita decidir se ele cumpre ou não uma certa e bem definida propriedade matemática, a partir de tal resultado se pode falar do argumento original em termos de correção/incorreção lógica. (Seoane 2014, p.184)

Mas como Seone mesmo afirma, esta estratégia nem sempre funciona e ela pode ser aplicada de modo débil quando a forçamos sobre argumentos que não permitem uma tal interpretação algorítmica. O que estes argumentos que são irredutíveis à estratégia de tradução nos indicam é que esta estratégia de tradução e cálculo não é o único modo pelo qual a lógica formal colabora com a compreensão de argumentos e que existem outras formas de interação entre lógica formal e os argumentos irredutíveis às formas dedutivamente válidas. Seone defende que há uma grande riqueza de interação entre lógica [formal] e argumentação: "as ferramentas conceituais da teoria lógica [formal] aparecem como um importante guia para resolver, de modo flexível, os problemas particulares de identificação e avaliação e como um importante guia para a reflexão geral a respeito da avaliação do trabalho justificador das premissas." (Seoane 2014, p.185) Assim, compreendemos que a lógica formal cumpre uma função fundamental no estudo e na construção de argumentos. Por outro lado, acreditamos também que é verdadeiro dizer que a lógica formal se beneficia das técnicas e métodos da lógica informal, ou seja, é verdadeiro dizer que a lógica informal complementa a lógica formal.

A lógica informal complementa a lógica formal pois para que ocorra o desenvolvimento da lógica formal, a construção de sistemas dedutivos, de novas lógicas, é preciso que haja um âmbito da experiência demonstrativa ainda

não formalizado. No contexto da construção ou do aspecto inventivo da lógica, é necessário que ela se debruce sobre aspectos da linguagem e do raciocínio que ainda não foram formalizados. Os esforços de categorização, sistematização e formalização da lógica informal permitem a compreensão das regras não escritas da argumentação. Para que o processo de formalização avance, como por exemplo vemos nos esforços da lógica deôntica, etc, a lógica informal e sua categorização e sistematização de proposições e argumentos fornece material para a lógica formal. Mas isto não quer dizer que a lógica informal só complementa a lógica formal na medida em que ela pode servir de material para formalização. A lógica informal e as análises dos princípios norteadores da argumentação em linguagem natural são importantes também para os debates a respeito dos critérios para avaliação dos sistemas formais bem como para as discussões relativas à natureza destes sistemas. Como nos lembra Kripke[22] em seu "elementary methodological sermon" de 1976:

> As investigações lógicas podem ser obviamente usadas como uma ferramenta útil da filosofia. Elas devem, entretanto, ser acompanhadas pela sensibilidade à significância filosófica do formalismo e por uma dose generosa de senso comum, assim como por um profundo conhecimento tanto dos conceitos básicos quando dos detalhes técnicos do material formal utilizado. Não se pode supor que o formalismo é capaz de forjar resultados filosóficos de uma maneira que ultrapasse o raciocínio filosófico mais básico (ordinary philosophical reasoning). Não existe substituto matemático para a filosofia[23] p. 416.

[22] Nestas páginas finais do artigo, Kripke organiza o debate, a crítica que recebeu e a crítica que realizou em termos de princípios gerais para debates em lógica formal. Estas páginas são preciosas pois ele oferece uma lista de princípios que chama de "elementary methodological morals", elencando os vícios de argumentação da literatura técnica. Estes princípios deveriam ser usados para nortear toda e qualquer discussão que envolva sistemas formais. Aqui os trago os princípios de Kripke como princípios gerais que devem nortear os esforços de definição da lógica informal e seus modos de interação com a lógica formal. Eles são os seguinte: 1) devemos tomar cuidado para não basear uma discussão filosófica em uma falsidade matemática ou em um trivialmente falso "*hard logical fact*"; 2) é fundamental termos em mente e realizarmos distinções conceituais básicas; 3) filósofos não devem confundir suas doutrinas filosóficas particulares com os resultados básicos e os procedimentos da lógica matemática; 4) a tendência em se propor critérios puramente técnicos com o propósito de se excluir estratégias que não gostamos deve ser reprovada. ("ao menos que à um critério puramente técnico tenha sido dada uma justificativa rigorosamente filosófica que realmente justifique que ele seja adotado à priori,…, é preciso que este critério seja considerado parte de uma hipótese com a qual o autor escolheu trabalhar"); 5) filósofos devem manter-se céticos frente à tentativas fáceis de se resolver questões linguísticas ou empíricas por meio de considerações formais a prior que sejam rápidas demais ("é certamente verdadeiro que alguns autores mais recentes de filosofia são tão apaixonados pela lógica formal que eles escrever como se ela pudesse ser utilizada para resolver questões admitidamente empíricas sobre o Inglês [linguagem natural] sem que se realize o mínimo de trabalho empírico"); 6) filósofos devem ter um senso melhor tanto do poder quanto das limitações das técnicas formais e matemáticas ("um pouco de senso comum é necessário para se manter sob controle a tendência dos filósofos de construir castelos formais imaginários [*formal castles in the air*]."

Essa afirmação de Kripke vai na direção da conclusão de Pereira no artigo "Brower e A Natureza da Lógica": é preciso não esquecer a máxima tautológica que diz "pensar é pensar, calcular é calcular" (Pereira 2008 p. 979). Confundir cálculo com o pensamento é um erro elementar. E porque partimos do princípio de que o pensar não se reduz ao cálculo, consideramos que a análise geral de argumentos (a lógica informal) nos permite deixar explícitos os princípios da lógica formal e servir como estratégia para o exercício da filosofia em geral e da filosofia da lógica.

Haack (2006) nos oferece mais evidências para o nosso argumento de que a lógica informal complementa a formal ao mostrar que a filosofia da lógica não pode ser plenamente substituída por um procedimento de cálculo sob pena de se perder um âmbito largo e rico de procedimentos de raciocínio que fazem parte do universo das experiências demonstrativas. Dado que a filosofia da lógica se ocupa de âmbitos como, por exemplo, a epistemologia (e a construção de lógicas epistêmicas) e a metafísica (e a identificação dos compromissos ontológicos de um sistema formal), que resistem à regimentação, seu exercício não poder ser realizado unicamente através de procedimentos matemáticos e não poder se reduzir às estratégias de tradução e cálculo[24]. Este fato evidencia que também que há uma diversidade rica de relações entre a lógica informal e a lógica formal.

Os modos de demonstração que nos servirão de parâmetro para nossa investigação sobre as relações entre lógica formal e informal são a dedução

---

[23] Saul Kripke (1976) "Is there a problem with substitutional quantification?" in Evans, G. & McDowell, J. (orgs.) *Truth and Meaning Essays in Semantics.* OUP, 1999. p.325-420

[24] Este mesmo problema se estende para a lógica, no contexto das discussões sobre demarcação já que são discutíveis os critérios para se determinar se um sistema formal é ou não uma lógica. Haack nos lembra que Dummet, por exemplo, argumentou em 1973 que as lógicas epistêmicas não são realmente lógicas "porque crença e conhecimento são noções irremediavelmente vagas" (Haack, Filosofias, p.33). Haack, entretanto, propõe que concedamos o benefício da dúvida a respeito do status como lógica de sistemas alternativos ao da lógica clássica. A vantagem do benefício da dúvida é que ela impede a abordagem excessivamente rigorosa e, por vezes dogmática, em relação aos sistemas alternativos. A análise de sistemas alternativos por analogia com a lógica clássica é um procedimento que dá à lógica clássica um *status* que é filosoficamente problemático, na medida em que ela é tomada como critério indiscutível e como modelo único para se definir o que é a lógica. Ainda que a lógica clássica seja um bom critério e um bom modelo do que deve ser considerado como uma lógica, por outro, este fato não pode levá-la a ser um critério indiscutível ou um que seja tomada como modelo único *a priori*. Para o benefício da lógica como uma prática, é necessário que seus critérios e modelos possam sempre passar por revisões racionais, nesse sentido é importante que o lugar da lógica clássica como fornecedora de critérios últimos seja colocado em questão.

natural e as suas regras [25], as provas por indução matemática e o método de análise de argumentos. Dentre os métodos formais de prova ou demonstração, iremos investigar em que medida os sistemas de dedução natural cumprem a motivação que se encontra em suas origens: servir como um sistema que simula as inferências realizadas naturalmente pelo pensamento de modo que suas regras parecem o menos artificial possível. Arthur (2011) nos lembra que "o método de dedução natural tem a vantagem de que está baseado em regras de inferência que refletem muito proximamente as formas com as quais raciocinamos intuitivamente, ou ao menos as regras que estão implícitas em nosso raciocínio" (p.17). Para analisar o caráter intuitivo dos sistemas de dedução natural devemos verificar em que medida as regra de introdução e de eliminação de conectivos segue o uso dos conectivos em linguagem natural e nos argumentos filosóficos na história da filosofia.[26]

Ortiz (2014), por exemplo, aponta o caráter contra-intuitivo da regra de introdução da disjunção, pois ela enfraquece a verdade de uma proposição ao adicionar a ela um disjunto falso. Consideremos o caso proposto por ele: "Antonio sabe hablar español". Se é verdadeira esta frase, temos que a frase "Antonio sabe hablar español o inglés" também é verdadeira. Entretanto, ao inserimos a disjunção para construir esta segunda frase, acabamos construindo uma afirmação mais fraca. Diferente de Ortiz, considero que há casos retóricos em que a introdução da disjunção funciona perfeitamente para reforçar a verdade de uma afirmação. E isto pode ser visto no exemplo que ele mesmo dá, mas que ele considera também um exemplo de enfraquecimento da afirmação original. O exemplo é o seguinte: "Madrid es la capital de España". Ortiz (2014) considera que ao adicionarmos um disjunto qualquer, esta afirmação torna-se mais fraca, inclusive o disjunto "yo soy Bruce Lee". É possível, entretanto, imaginar um caso onde dois interlocutores estão conversando sobre capitais dos países da Europa, nenhum dos dois é, tem o mesmo nome ou se parece com Bruce Lee, mas ambos são fãs do ator. Se neste caso, um

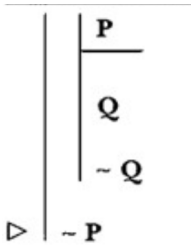
[25] Os sistemas de dedução natural que nos interessa aqui são os chamados SD, SD +, PD, PD +, PDE

[26] Nossa análise tem como inspiração de base o puzzle lógico de Peter Wason (1966), o chamado *Wason selection task*. Wason construiu um puzzle lógico com o qual conseguia avaliar em que medida o uso intuitivo do condicional seguia, acompanhava, o seu uso formal. Outras referências para o tipo de análise do uso de conectivos que apresento aqui são: Sautter (2015) em seu "Compreensão lógica e compreensão cotidiana" e Ortiz (2014) em "Um sistema natural de deducción natural".

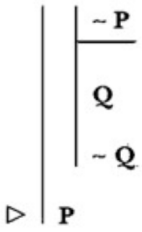
deles afirmar que Madri é a capital da Espanha e outro discordar, dizendo que na verdade a capital da Espanha é Barcelona, então, o primeiro poderá, em um movimento retórico bem sucedido, introduzir a disjunção para fins de convencimento. Neste caso, quando ele enuncia: "Madri é a capital da Espanha ou Eu sou o Bruce Lee", a verdade da afirmação original fica mais evidente e mais forte. Assim, podemos ver que, em certos casos, utilizamos sim a introdução da disjunção para facilitar o convencimento.

Outra vantagem de tomar os sistemas de dedução natural como fonte do estudo das noções de inferência, e demonstração, capazes de nos informar sobre as relações entre lógica formal e informal, entre filosófica e lógica, é que este sistema opera com regras que podem ser consideradas elas mesmas tipos específicos de prova. Vejamos alguns exemplos:

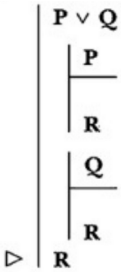
Regra de introdução da negação:



Regra de eliminação da negação:



Regra de eliminação da disjunção:



A regra da introdução e da eliminação da negação são, em si mesmas, provas por contradição (ou *reductio ad absurdum*), ou seja, assumimos que a negação do que queremos demonstrar é verdadeiro. Dela derivamos uma contradição e, portanto, demonstramos que a hipótese de que a negação do que queríamos demonstrar é falsa. A eliminação da disjunção, por sua vez, pode ser apresentada como uma prova por análise de casos. Avaliamos se com o disjunto esquerdo podemos derivar uma sentença R qualquer. Depois avaliamos se com o disjunto direto podemos também derivar R. Se for o caso que de ambos os disjuntos podemos derivar R, então, podemos derivar R da disjunção inicial. Todas as derivações nos sistemas de dedução natural funcionam como a regra de introdução do condicional, elas são um raciocínio hipotético no qual partimos das premissas ou dos axiomas e verificamos se podemos chegar à fórmula que tomamos como conclusão. Com a análise destes exemplos prosseguiremos a investigação sobre as relações entre regras de inferência formais e raciocínio informal.

O segundo modo de demonstração que tomaremos como objetos de estudo a fim de analisar as diferenças e semelhanças entre a lógica formal e a lógica informal são as provas por indução matemática como elas aparecem na lógica. Enquanto a dedução natural é uma técnica de manipulação simbólica para realizar demonstrações no interior de um sistema dedutivo, a indução matemática é usada para demonstrar resultados sobre a semântica e a sintática destes sistemas. Por isso, a indução matemática é um modo de demonstração comumente utilizado na metateoria dos sistemas lógicos. O método de indução matemática é uma estratégia que faz uso de definições recursivas e de um

passo indutivo que, conjuntamente, permitem demonstrar que um resultado vale para um número infinito de itens. Ele funciona para demonstrar metateoremas nos sistemas lógicos porque não só as sentenças formadas são compostos verofuncionais definidos recursivamente como sua construção é fundada no princípio de composicionalidade. Ainda que a indução matemática funcione pela introdução do passo indutivo, não podemos considerá-la como idêntica à a indução empírica. Como diz Max Back (1971): "a indução matemática não é uma forma especial da indução ordinária; ela é uma variedade de prova rigorosa: uma demonstração ou dedução"[27]. Assim, consideramos que as provas por indução matemática constituem um objeto importante para a presente investigação. Tomemos, como exemplo para nossa explicação metodológica, a prova do lema TFI-1 da semântica da lógica sentencial[28]:

> (*Lema TFI-1):* Se *P* é uma sentença em LS sem nenhum conectivo binário, então *P* é vero-funcionalmente indeterminado.

A prova deste lema, que é realizada considerando o tamanho da sentença *P*, contém uma cláusula base e um passo indutivo. Vejamos:

> *Prova:* por indução no comprimento de *P*
> *Passo Base:* Suponha que *P* seja uma sentença negada de forma ~*P* de tamanho zero. Então, *P* é uma sentença atômica, ou seja, *P* é uma letra sentencial. Agora defina as funções $\mathfrak{U}_1$ e $\mathfrak{U}_2$ de atribuição de valor de verdade do seguinte modo:
>
> $\mathfrak{U}_1(R) = T$ para todas as letras sentenciais R
> $\mathfrak{U}_2(R) = F$ para todas as letras sentenciais R

[27] Black, Max em Encyclopedia Americana, vol. 15, 1971. p.100 cf.. Carnielle & Epstein. "Computabilidade, funções computáveis, lógica e os fundamentos da matemática" Editora UNESP, São Paulo. 2005 p.59.

[28] A prova deste lema, bem como de outros resultados em LS e LP utilizados aqui foram tiradas de Klenk (2002) e Bergman, Moore & Nelson (2009).

Ou seja, para cada letra sentencial, podemos aplicar a função $\mathfrak{A}_1$(R) ou $\mathfrak{A}_2$(R). Portanto, $\mathfrak{A}_1$(*P*) é verdadeiro e *P* é verdadeiro em $\mathfrak{A}_1$ e $\mathfrak{A}_2$(*P*) é falso e *P* é falso em $\mathfrak{A}_2$. Logo, *P* é vero-funcionalmente indeterminado.

Passo indutivo: Suponha que TFI-1 seja verdadeiro para todo *P* de tamanho K ou de tamanho menor que K. Assuma que *P* é uma sentença negada de forma ~*P* de tamanho K + 1. Sendo assim, *P* é uma sentença de forma ~ Q, onde Q é sentença negada de tamanho K. Por nossa hipótese indutiva, Q é vero-funcionalmente indeterminado. Deste modo, há funções de atribuição de valor de verdade $\mathfrak{A}_3$ e $\mathfrak{A}_4$ de modo que Q é verdadeiro em $\mathfrak{A}_3$ e falso em $\mathfrak{A}_4$. Assim, ~Q, ou seja, P, é falso em $\mathfrak{A}_3$ e verdadeiro em $\mathfrak{A}_4$. Portanto, se *P* é uma sentença de qualquer tamanho, seja 0, seja K, ou K +1, dado que P seja uma sentença sem nenhum conectivo binário, então P é vero-funcionalmente indeterminado.

Esta prova mostra que quando queremos mostrar teses sobre a semântica e a sintaxe de sistemas como a lógica sentencial ou a lógica de predicados de primeira ordem, realizamos raciocínios em linguagem natural. Entretanto, estas provas em linguagem natural tem uma estrutura argumentativa bem formada que é, de certo modo, uma estrutura padrão. Para que a lógica seja analisada sob seu aspecto metodológico, é importante mostrar que a lógica, em um sentido mais largo, não tratar apenas de demonstrações em linguagem simbólica, mas inclui também demonstrações em linguagem natural. Essa é a importância de passarmos pela análise das provas por indução matemática em nosso projeto. Além da prova por indução matemática, outro modo de demonstração que faz uso da linguagem natural é o o método de análise de argumentos. O método de análise de argumentos, no entanto, não faz parte dos procedimentos da lógica formal, mas da prática informal. No método de análise de argumentos os passos de demonstração não são regimentados. Entretanto, os argumentos quando categorizados pelo método de análise, podem ser interpretados a partir de sua semelhança estrutural. O método da análise de

argumentos é realizado a partir de certos princípios de reconstrução e avaliação. Esta estratégia de análise é utilizada não apenas nas aulas de lógica informal, mas também na construção de argumentos filosóficos, ou seja, ela é recorrente tanto no ensino quanto na pesquisa. A análise de argumentos é uma estratégia *formalmente muito simples*. Ela é constituída de duas partes: a reconstrução e a avaliação.

A análise ocorre de modo que todo e qualquer argumento, antes de ser avaliado, deve ser reconstruído e colocado em sua forma padrão. A reconstrução de argumento é feita de modo a excluir da avaliação qualquer ruído ou ambiguidade. Na reconstrução também adiciona-se de modo a tornar explícita qualquer premissa fundamental que o autor do argumento tenha deixado implícita. Para tanto, é preciso que se mapeie as sentenças a partir das relações que mantém entre si. De um modo geral, o que isso significa é que diferenciamos premissas, conclusões intermediárias e conclusão final. As premissas são reescritas e encadeadas de modo que o argumento reconstruído em sua forma padrão esteja bem formado. A vantagem da reconstrução é que ela torna argumentos com estrutura complexa mais claros. A simplicidade do método de reconstrução, entretanto, não indica que ele é trivial. O esforço da reconstrução é em deixar claro o papel de cada proposição e sua forma (universal, particular, afirmativa ou negativa), como um passo (análogo à tradução) e que é preparatório para a avaliação (nesse caso, um passo que é análogo ao cálculo).

Há dois princípios que guiam a reconstrução de argumentos: o princípio de caridade e o princípio de fidelidade. Estes princípios são importantes de serem explicitados porque a razão para se reconstruir um argumento não é exatamente refutá-lo ou defendê-lo, mas tornar sua estrutura explícita para melhor compreendê-lo. O propósito da reconstrução é, portanto, produzir um "conjunto de sentenças que representem o argumento real - ou seja, as premissas, a conclusão, e a relação entre as premissas e a conclusão - que o autor procurou apresentar." (Hughes 2015, p.67) Nesta reconstrução, revisamos os termos utilizados pelo autor e os rescrevemos de modo a eliminar ambiguidades e construir um argumento padrão bem formado. Face a uma ambiguidade ou falta de clareza, ao invés de reconstruirmos o argumento de modo que ele

fique fácil de ser refutado, o tornamos mais claro e mais forte. Assim, o princípio de caridade é importante porque nos força a analisar o argumento em sua forma mais razoável, mais forte possível. A reconstrução caridosa, entretanto, pode levar a excessos. Assim, devemos também atentar para o princípio de fidelidade. Não se deve reconstruir um argumento oferecendo uma versão padrão que viole a intenção do autor (ainda que esta não esteja clara). O argumento reconstruído deve ser consistentes com as intenções do autor e, se adicionarmos premissas que consideramos estar implícitas, estas devem ser razoáveis e não obviamente falsas. Neste momento da análise há também o perigo de se reconstruir o argumento de modo a termos uma validade barata ("*cheap validity*"), ou seja, ganharmos de modo fácil, por estarmos encarregados da reconstrução das premissas, um argumento dedutivamente válido. Esse é um dos problemas da abordagem dedutivista, pois não é o caso que a reconstrução do argumento como um argumento dedutivamente válido sempre nos dará a versão mais forte e persuasiva do argumento.

A avaliação do argumento exige duas coisas: o mapeamento do tipo de argumento em questão e a crítica informada (teoricamente embasada) das premissas. Falemos primeiro dos princípios e depois dos tipos gerais de argumento (tipificados por Feldman 1999, Morrow & Weston 2011, Bonevac 1999). Os princípios que regem a avaliação de argumentos guiam toda e qualquer boa crítica filosófica ou científica. São princípios tão gerais e bem estabelecidos que beiram a trivialidade. Estes princípios na verdade são regras gerais que sistematizam toda boa contra-argumentação. Por exemplo: negar a verdade (e demonstrar a falsidade) da conclusão não é aceitável do mesmo modo que aceitar um argumento porque a conclusão é verdadeira não é. A contra-argumentação deve ser direcionada a premissas individuais e elas devem ser criticadas de modo específico, a partir de razões substanciais e relevantes para a questão.

A partir deste método que busca encontrar as estruturas gerais dos argumentos e oferecer, posteriormente, um conjunto de regras gerais ou princípios para avaliação deles, foram categorizados diferentes tipos de argumento. Para cada um destes argumentos há uma forma padrão. Estas formas padrão

podem ser construídas como estruturas de argumentos dedutivos ou não-dedutivos, dependendo do contexto. De todo modo, não há um procedimento mecânico e único para se acessar a estrutura do argumento e muito menos para avaliar a força indutiva de argumentos não-dedutivos. Ou seja, em argumentos não-dedutivos, não há exatamente como medir ou demonstrar qual o grau de certeza com a qual a conclusão se segue das premissas.

O método de análise de argumento não é um procedimento mecânico de tradução ou de cálculo e suas regras pouco tem em comum com as regras de demonstração matemáticas. Por este motivo ele é, dentre os três métodos acima descritos, o que mais se assemelha ao processo pelo qual a análise filosófica ocorre no âmbito da pesquisa contemporânea. Compreender e analisar obras de grandes autores da história da filosofia exige do pesquisador a capacidade de reconstruir os argumentos principais e formular críticas, objeções e defesas. A parte do nosso posicionamento pessoal frente aos textos clássicos é uma dimensão da análise que pode ser facilmente deixada de lado ou feita de modo pouco estruturado. Por isso o estudo da análise de argumentos se mostra fundamental para o exercício da argumentação filosófica. O método de análise de argumentos nos permite reconstruir os argumentos de obras fundamentais na história da filosofia e nos auxilia para a organização sistemática dos tipos de demonstração e os modos como os grandes autores e autoras realizaram sua argumentação filosófica. Entendo que este artigo abre inúmeras questões para serem aprofundadas mais detidamente em outros trabalhos. Mas depois de ter feito uma leitura de sobre-vôo, apresentando criticamente o uso e o fim dos métodos de demonstração informais e formais, penso ter mostrado tanto o papel metodológico da lógica na filosofia como a continuidade entre as formas de demonstração neste contexto.

Referências

Aristotle. *The Complete Works of Aristotle.* Jonathan Barnes (ed.) 2 Volumes. Princeton U Press, 1941.

Arthur, R. T. W. *Natural Deduction An Introduction to Logic with Real Arguments, a Little History, and Some Humour.* Broadview Press, 2011.

Bergamnn, Moor & Nelson.*The Logic Book.* 6th Edition. McGrawHill, 2014.

Bishop-Clark & Dietz-Uhler. *Engaging in the Scholarship of Teaching and Learning: a guide to the process, and how to develop a project from start to finish.* Stylus Publishing, 2012.

Blanché, R. & Dubucs, J. *História da Lógica*. Edições 70, 1996.

Bonevac, D. *Simple Logic.* Oxford University Press, 1999.

Boole, G. *An Investigation of the Laws of Thought on which are founded the mathematical theories of logic and probabilities*. Dover Publications, 1958.

Brunschvicg, L. *Les Etapes de La Philosophie Mathematique*. University of Michigan Historical Reprint Series- U Michigan Press, 1922.

Cáceres, D.S. G " La lógica informal como eje para la enseñanza del Taller de Lectura, Reddacción e Iniciación a la investigación documental III" em *Revista de Filosofía Universidad Veracruzana.* N. 30-21 México. Set 2014 -Marzo 2015 pp. 97-107

Carnielle & Epstein. "Computabilidade, funções computáveis, lógica e os fundamentos da matemática" Editora Unesp: 2005 p.59.

Casanave, A. L. (ed.) *Symbolic Knowledge from Leibniz to Husserl.* College Publications, 2012.

Copi, I. & Gould, J. (orgs.) *Contemporary Readings in Logical Theory.* Macmillan, 1967.

Corcoran, John. "Critical Thinking and Pedagogical License". *Manuscrito* XXII, 109-116. (1999)

_____. "Logic Teaching in the 21st Century." em *Revista de Argumentación y Retórica*. Universidad de Guadalajara Vol1.N1. 2016

Cozic, M. & Bonnay, D. (orgs.) *Philosophie de la Logique. Conséquence, preuve et vérité.* Vrin, 2009.

Dos Santos, Luiz Henrique Lopes. *O Olho e o Microscópio.* Nau Editora, 2008.

Esquisabel, O. "Representing and Abstracting: An analysis of Leibniz's Concept of Symbolic Knowledge" em Casanave, A.L. (ed.) *Symbolic Knowledge from Leibniz to Husserl*. College Publications, 2012.

Ewald, William Bragg. *From Kant to Hilbert.* Volume 1. Oxford University Press, 2007.

Feldman, R. *Reason and Argument.* Prentice Hall, 1999.

_____. "Thinking, Reasoning, and Education," in Harvey Siegel (ed.), *Oxford Handbook of Philosophy of Education*, Oxford: Oxford University Press, 2009.
Frege, G. *Lógica e Filosofia da Linguagem*. Edusp, 2009.

Grattan-Guinness. "The Mathematical Turn in Logic" in *The Handbook of the History of Logic. Volume 3: The Rise of Modern Logic from Leibniz to Frege.* Ed. Gabbay & Woods. Amsterdam: Elsevier, 2004.

Grennan, W. *Informal Logic: Issues and Techniques: A proposal for a new system of argument evaluation.* McGill-Queen's University Press, 1997.

Groarke, Leo, "Informal Logic", *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2013/entries/logic-informal/>.

Guttenplan, S. *The Languages of Logic.* Blackwell, 1997.

Haack, S. *Filosofias das Lógicas*. Editora Unesp, 1998.

Halper, E. "Aristotle's Scientific Method" in Wians & Polansky (orgs.) *Reading Aristotle Argument and Exposition.* Brill, 2017.

Hansen, Hans V.,. "Are there methods of informal logic?" in F. Zenker (ed.), *Argumentation: Community and Cognition, Proceedings of the 9th Conference of the Ontario Society for the Study of Argumentation*, Windsor: CRRAR, 2011.

Hintikka, Jaakko. "The Role of Logic in Argumentation" in *The Monist*, Vol. 72, No. 1, Logic as a Field of Knowledge (Jan 1989), pp. 3-24

_____. *Knowledge and Belief. An Introduction to the Logic of the Two Notions.* Cornell University Press, 1962.

Hodges, W. *Logic an introduction to elementary logic*. Penguin Books, 2001.

Hughes, Lavery & Doran. *Critical Thinking An Introduction to the Basic Skills.* Broadview Press, 2015.

Ishiguro, H. Leibniz's Philosophy of Logic and Language. Cambridge University Press, 1991

Johnson, R. H. "Informal Logic and Deductivism," *Studies in Logic*, 4(1): 17–37., 2011

_____. *The Rise of Informal Logic*, Newport News, VA: Vale Press (1996); Digital edition, University of Windsor: Windsor Studies in Argumentation, 2014.

Kneale & Kneale. *The Development of Logic.* Clarendon Press, 2008.

Klenk, Virginia. *Understanding Symbolic Logic.* Prentice Hall, 2002.

Kripke, S. (1976) "Is there a problem with substitutional quantification?" in Gareth Evans & John McDowell (orgs.) *Truth and Meaning Essays in Semantics.* (1999) OUP p.325-420

_____. *Naming and Necessity*. Blackwell, 1972.

Legris, J. "Between Calculus and Semantic Analysis. Symbolic Knowledge in the Origins of Mathematical Logic" em Casanave, A.L. (ed.) *Symbolic Knowledge from Leibniz to Husserl*. College Publications, 2012.

Leitão, S. "O lugar da argumentação na construção do conhecimento em sala de aula." In: Leitão, S.; Damianovic, M. C.a (Org.). *Argumentação na escola*: *o conhecimento em construção*.Pontes Editores, 2011, p. 13-46.

Luque, L. "Logic as (Normative) Inference Theory: Formal vs Non-formal theories of inference goodness." in *Informal Logic*, 28 (4): 315-334.2009

MacFarlane, J. G. *What does it mean to say that logic is formal?* 328 f. Tese (Doutorado em Filosofia). University of Pittsburgh, 2000.

Malpass & Marfori. *The History of Philosophical and Formal Logic. From Aristotle to Tarski.* Bloomsbury Academic, 2017.

Mancosu, P. *From Brouwer to Hilbert.* Oxford University Press,1998.

_____. *Philosophy of Mathematics and Mathematical Practice in the Seventeenth Century.* Oxford University Press, 1996.

Maudet, Nicolas & Fisher, Alec. "Informal Logic and its Implications for Philosophy." in *Informal Logic*. 20 (2).2000.

Mortari, C. *Introdução à lógica*. Ed. Unesp, 2001

Novaes, C. D. *Formal Languages in Logic. A Philosophical and Cognitive Analysis.* Cambridge University Press, 2012.

Ortiz, H. H. "Un sistema natural de deducción natural" em *Revista de Filosofía Universidad Veracruzana.* N. 28-29. Set 2013 - Marzo 2014. pp. 23-43

Patterson, R. *Aristotle's Modal Logic. Essence and Entailment in the Organon.* Cambridge University Press, 1995.

Pereira, L. C. P. D. & Wrigley, M. (Org.). *Logic, Language and Knowledge: Essays in Honour of Oswaldo Chateaubriand Filho.* CLE- UNICAMP, 1999. v. xxii. 573p .

Pereira, L. C. P. D. " A Natureza da Logica" em *Anales de la Academia de Ciencias de Buenos Aires*. 1ed. Buenos Aires: Academia Nacional de Ciencias de Buenos Aires, 2008, v. XLI, p. 969-980.

Prawitz D. "Remarks on some approaches to the concept of logical consequence," *Synthese*, 62: 153–171. 1985.

_____. *Natural Deduction A Proof-Theoretical Study.* Dover Publications, 2006.

Sautter, F. T. "Compreensão lógica e compreensão cotidiana" em Secco, G. (org.) *Epistemologia e Currículo: registros do II Workshop de Filosofia e Ensino*. ED.UFRGS, 2016.

Secco, G. & Pugliese, N. "On How Logic is presented to the Brazilian Student: a critical Analysis" in Teresita de Jesús Mijangos Martínez (Org.). em *Anales del Encuentro de Didactica de Lógica - Academia Mexicana de Lógica* A.C., V.1, 2016. pp.78-92

Secco, G. D. "Filosofia no Ensino Médio: distinções preliminares para uma didática mínima da lógica" em *Controvérsia*, V.09, pp.89-102, 2013.

Tarski, Alfred, 1986, "What are Logical Notions," *History and Philosophy of Logic*, 7: 143–154.

Velasco, Patricia. "Sobre o lugar da argumentação na filosofia como disciplina" em *Educação e Filosofi*a, v. 31, n.61, 2017.

_____. "Argumentação, Pensamento Crítico e Ensino de Filosofia: (im)precisões conceituais." In: Carvalho, Marcelo; Benedito de Almeida Jr., J. *Filosofia e Ensinar Filosofia*. Coleção XVI Encontro ANPOF: ANPOF, 2015, p. 224-236.

_____. *Educando para a Argumentação*: *contribuições do ensino da lógica*. Autêntica Editora, 2010.

_____. "Sobre o Lugar da Lógica na Sala de Aula" em *Revista Sul-Americana de Filosofia e Educação*. n.13, 2009.

Villareal, A. C. "Competencia en didáctica de la lógica" em *Revista de Filosofía Universidad Veracruzana.* N. 17 México. Set. 2005. pp. 37-47

Wohlrapp, Harald R. *The Concept of Argument. A philosophical foundation.* Dordrecht: Springer. 2014.

Woods, J. "The Necessity of Formalism in Informal Logic" in *Argumentation* 3, 1989.

FILOSOFIA DA LÓGICA EMPIRICAMENTE INFORMADA: UMA ILUSTRAÇÃO VIA METAFÍSICA DA LÓGICA

César Frederico dos Santos[1]

RESUMO

Dados da psicologia são relevantes na abordagem de alguns problemas em filosofia da lógica mesmo que não se assuma, de antemão, uma posição psicologista. Para ilustrar um método de abordar problemas na filosofia da lógica que faz largo uso de dados da psicologia, neste artigo considero a questão sobre a metafísica da lógica—qual o objeto de estudo da lógica?—à luz de resultados da psicologia do pensamento e da psicologia do desenvolvimento. Estes resultados nos permitem concluir que sistemas lógicos não são meramente descritivos de aspectos gerais da linguagem natural, do mundo ou da racionalidade humana. Os dados são consistentes com a hipótese de que sistemas lógicos dedutivos sejam conjuntos de regras de caráter primordialmente não-descritivo, criadas mediante reflexão ativa de filósofos, matemáticos e lógicos, visando garantir inferências dedutivas em certos contextos.

Palavras-chave: Filosofia da Lógica. Filosofia Empiricamente Informada. Psicologia do Pensamento. Psicologia do Desenvolvimento. Metafísica da Lógica.

ABSTRACT

Data from psychology is important to address some questions in philosophy of logic, even if one does not assume, from the outset, a psychologistic stance. In order to illustrate a method of approaching problems in philosophy of logic that makes extensive use of psychology data, in this paper I consider the question of the metaphysics of logic—what is logic about?—in the light of findings in psychology of reasoning and developmental psychology. These findings allow us to conclude that deductive logical systems are not merely descriptive of structural aspects of natural language, the world, or human reasoning. The data is consistent with the hypothesis that logical systems are sets of primarily non-descriptive rules, created by philosophers, mathematicians, and logicians by means of active reflection in order to ensure deductive inferences in certain contexts.

Keywords: Philosophy of Logic. Empirically-Informed Philosophy. Psychology of Reasoning. Developmental Psychology. Metaphysics of Logic.

[1] Doutorando na Vrije Universiteit Amsterdam. Professor assistente da UFMA. Bolsista FAPEMA.
E-mails: c.f.dos.santos@vu.nl / cesar.frederico@ufma.br. ORCID: 0000-0003-4458-5250.

1. Introdução

A proposta deste artigo, como seu título indica, é mais ilustrar um método de abordar problemas na filosofia da lógica do que propriamente responder a problemas da filosofia da lógica. O método que quero ilustrar atende por vários nomes, cada um correspondendo a uma nuance sua—naturalismo, filosofia segunda, filosofia experimental, filosofia empiricamente informada, filosofia sintética—, mas basicamente consiste no uso de dados empíricos, coletados de maneira científica, para informar a abordagem de questões filosóficas. A questão da filosofia da lógica que selecionei para exemplificar a aplicação deste método é a investigação sobre o objeto de estudo da lógica. Esta investigação é por vezes chamada de "metafísica da lógica", como fazem Cohnitz & Estrada-González (2019), por exemplo, em seu recentemente publicado manual introdutório à filosofia da lógica. Parece um contrassenso supor que podemos investigar questões *meta*físicas usando dados empíricos que, obviamente, dizem respeito somente ao mundo *físico*. Contudo, um olhar mais atento às questões investigadas pela assim chamada metafísica da lógica logo mostra que é viável abordar estas questões sob uma perspectiva empiricamente informada. Cohnitz & Estrada-González, por exemplo, desdobram a questão da metafísica da lógica nas seguintes sub-questões:

> qual é o objeto que investigamos quando estudamos lógica? O objeto de estudo da lógica seria um conjunto de fatos muito gerais sobre o mundo (…), ou seria a lógica o estudo de fatos gerais sobre nossas mentes? (Afinal de contas, diz-se que a lógica estabelece as "leis do pensamento.) Ou seria a lógica apenas o estudo de convenções linguísticas (por exemplo, sobre como usar 'ou' e 'não')—sendo que talvez estas convenções sejam introduzidas e moldadas em parte pelo próprio estudo da lógica? (Cohnitz & Estrada-González, p. 113, minha tradução).

Como fica claro, estas perguntas tratam da conexão entre a lógica e outros fenômenos: o mundo físico, a mente humana e a linguagem, todos eles passíveis de investigação empírica. O que é preciso mostrar é como dados empíricos sobre o mundo físico, a mente humana e a linguagem lançam luz sobre a pergunta filosófica pela "metafísica" da lógica. Isto é o que farei aqui.

A pergunta sobre o objeto de estuda da lógica é uma questão bastante ampla. Talvez o correto fosse perguntar pelo*s* objeto*s* de estudo da*s* lógica*s*, visto que há inúmeros sistemas de lógica, e não podemos afastar de antemão a possibilidade de que diferentes lógicas sejam sobre diferentes fenômenos. Contudo, para tornar a questão mais tratável no espaço de um artigo, aqui vou me restringir a lógicas dedutivas—isto é, aquelas em que a noção de consequência lógica é tal que a verdade das premissas garante necessariamente a verdade da conclusão—que quantificam, no mínimo, sobre objetos, e que têm alguma pretensão de aplicabilidade a inferências sobre o mundo físico. Há sistemas lógicos que são teorias matemáticas "puras", não aplicadas, e é possível aplicar sistemas lógicos a contextos não-discursivos, como a circuitos elétricos, por exemplo. Sobre estes sistemas e estas aplicações não tenho o que dizer aqui.

A melhor maneira de apresentar um método é aplicá-lo, isto é, mostrá-lo operando. Antes disso, porém, é conveniente ter uma breve visão geral do método. Isto é o que apresento na seção seguinte. Em seguida, na seção 3, mostro como podemos derivar consequências empíricas testáveis de concepções metafísicas da lógica que a veem com capturando aspectos do mundo físico, da racionalidade humana e da linguagem. Feito isso, nas seções 4, 5 e 6 mostro o método em operação: coloco essas concepções "metafísicas" à prova contra resultados experimentais da psicologia. Como veremos, esses resultados experimentais não sustentam integralmente as hipóteses de que a lógica seja sobre o mundo físico, sobre a racionalidade humana e sobre a linguagem. A conclusão sugerida, na seção 7, é que sistemas lógicos sejam *criações* de caráter não-descritivo que estabelecem ou sistematizam técnicas de inferência criadas por lógicos, filósofos e matemáticos para lidar com certos contextos de investigação científica.

2. O método

Como dito acima, a prática de usar dados científicos para informar a abordagem de problemas filosóficos responde por vários nomes: naturalismo,

filosofia segunda, filosofia experimental, filosofia empiricamente informada, filosofia sintética. Vejamos rapidamente cada uma dessas denominações.

O nome 'naturalismo' remete diretamente a Quine, que via a filosofia como contínua com a ciência, e os métodos científicos como tipicamente mais confiáveis do que os métodos da assim chamada filosofia primeira, isto é, aquela filosofia que pretende ser o fundamento da ciência (Hylton, 2007). O espírito do que vou apresentar aqui está, no geral, alinhado com essa visão quiniana. Em especial, não há dúvida de que os métodos científicos sejam os mais adequados para tratar de questões *empíricas*. Assim, se um problema filosófico está relacionado a questões empíricas, não há dúvida de que uma abordagem científica não pode ser dispensada. A questão fundamental é determinar se um dado problema filosófico é empírico ou tem uma versão empírica; nesses casos, a abordagem naturalista se impõe.

'Filosofia segunda' é o nome que Maddy dá à sua variante do naturalismo quiniano. Para Maddy, a filosofia vem somente depois da ciência, como uma atividade de reflexão crítica sobre os resultados desta última. A obra homônima de Maddy, *Second Philosophy* (2007), e em especial seu tratamento da filosofia da lógica, é uma importante fonte de inspiração para o que vou apresentar aqui, como teremos a oportunidade de ver.

Ao passo que tanto Quine quanto Maddy usam resultados científicos (Quine nem tanto) mas não os produzem eles mesmos, na filosofia experimental são os próprios filósofos que "põem a mão na massa", por assim dizer, e fazem pesquisa empírica seguindo os protocolos científicos relevantes, a fim de levantar dados que vão subsidiar suas investigações filosóficas. Do campo da filosofia experimental, a principal influência aqui é Machery (2017) e sua proposta de reformulação do método de análise conceitual. Para Machery, conceitos são entidades psicológicas, "corpos de informação [que as pessoas têm] sobre indivíduos, classes, substâncias ou eventos" (Machery, 2017, p. 210, minha tradução). Em sua abordagem, o tradicional método de análise conceitual se converte em uma investigação psicológica das crenças e comportamentos das pessoas com respeito ao conceito sob investigação. Assim, não é difícil ver que vários problemas típicos da filosofia analítica ganham uma con-

traparte empírica, que pode se beneficiar de pesquisa científica. Isso aparecerá na abordagem do conceito de objeto, mais abaixo. Uma observação importante: neste texto, sempre uso a palavra 'conceito' no sentido definido por Machery, isto é, para designar "corpos de informação" armazenados nas mentes de indivíduos humanos, e não entidades abstratas. Assim, "conceito de objeto" é aqui entendido como aquilo que indivíduos pensam sobre objetos.

'Filosofia empiricamente informada' é a expressão que Dutilh Novaes usa para caracterizar um dos vários métodos que ela emprega em Dutilh Novaes (2012). Como a expressão sugere, nesta abordagem as respostas a problemas filosóficos são informadas por resultados empíricos, tal qual nas várias formas de naturalismo. Contudo, diferentemente do naturalista quiniano, que concede à ciência a palavra final, o filósofo empiricamente informado se reserva o direito de não tomar as teorias científicas pelo seu valor de face.

> [É] perfeitamente possível que os filósofos se envolvam de forma crítica com o material empírico (em particular, questionando suposições e pressupostos ocultos), ao mesmo tempo que reconhecem o seu valor para a teorização filosófica. Com efeito, muitas vezes é possível "filtrar" estas análises de modo a separar a confusão conceitual de resultados empíricos seguros (Dutilh Novaes, 2012, p. 253, minha tradução).

O que é particularmente interessante no modo como Dutilh Novaes caracteriza seu modo de fazer filosofia empiricamente informada é que o direito do filósofo de avaliar criticamente os resultados científicos e apontar confusões conceituais não advém de um suposto acesso privilegiado a uma filosofia primeira, a partir da qual o filósofo criticaria a ciência. Na verdade, a crítica filosófica é feita também em bases científicas. Uma *expertise* tipicamente filosófica, que pode faltar ao cientista, é a capacidade analítica, reforçada por ferramentas lógico-formais e matemáticas. Outro ponto em que o filósofo pode contribuir advém da interdisciplinaridade e transversalidade que caracterizam o interesse filosófico. O trabalho científico contemporâneo é caracterizado pela hiperespecialização, de sorte que um cientista às vezes não conhece campos de pesquisa mais distantes que poderiam ter algo a dizer sobre o fenômeno que ele estuda. O filósofo, com sua visão mais distante e mais abrangente, está na posição ideal para fazer a junção de peças do quebra-cabeças

científico que os cientistas mesmos ainda não fizeram. É aí que entra a assim chamada "filosofia sintética." O termo, cunhado por Schliesser (2019), caracteriza

> um estilo de filosofia que reúne insights, conhecimentos e argumentos das ciências especiais com o objetivo de oferecer um relato coerente de sistemas complexos e conectá-los a uma cultura mais ampla ou a outros projetos filosóficos (ou ambos) (Schliesser, 2019, p. 1-2, minha tradução).

Basicamente, a proposta da filosofia sintética é reunir toda a informação relevante para responder a uma determinada questão e que está espalhada por diversas disciplinas que, na maioria das vezes, têm pouca comunicação entre si. Daí o nome "sintética", que não é empregado aqui em oposição à "analítica". Uma abordagem sintética também se beneficia da análise conceitual, tanto no sentido definido por Machery—análise psicológica de conceitos mentais—, quanto no que tange à reengenharia conceitual usando a caixa de ferramentas típica da filosofia analítica.

Vejamos como esses métodos podem ser empregados para abordar a questão em tela.

3. Convertendo metafísica em questões tratáveis empiricamente

A tradição brasileira em filosofia da lógica, especialmente aquela situada em torno do pensamento de Newton da Costa, tem no recurso à ciência uma parte central do seu método de pesquisa. A razão disso está na visão de da Costa a respeito do lugar onde lógicas—os objetos de estudo da filosofia da lógica—podem ser encontradas. De acordo com o que da Costa chama de "princípio da sistematização", "nos contextos racionais encontra-se, sempre, de modo explícito ou implícito, um sistema lógico" (da Costa, 1994, p. 45). A tarefa do lógico ou do filósofo da lógica é sistematizar os contextos racionais de interesse, explicitando suas lógicas subjacentes por meio de sistemas axiomáticos formais. Visto que a ciência é considerada o paradigma da racionalidade, as teorias científicas, especialmente as da física, têm sido o objeto de estudo preferencial dos filósofos desta tradição. A idéia é que a explicitação das lógicas subjacentes a teorias científicas poderia ajudar a responder certos

problemas em filosofia da lógica como, por exemplo, se existem contradições verdadeiras sobre o mundo (Arenhart, 2018), ou se o conceito de identidade tem aplicação universal (Krause & Arenhart, 2019).

Esta é, sem dúvida, uma forma bastante produtiva de abordar problemas filosóficos via resultados científicos. Mas é possível avançar no uso de dados da ciência na filosofia da lógica também em outra direção. Os dados que a ciência pode fornecer para a filosofia da lógica vão muito além de teorias racionais aptas à formalização.

Há um grupo de questões filosóficas ligadas à lógica, que poderíamos chamar de metafísicas, que diz respeito à relação entre lógica e mundo. Para investigar esse tipo de questão, precisamos tanto de dados sobre o mundo físico, quanto de dados sobre o mundo humano. A necessidade de dados do mundo físico dispensa explicações. Mas a necessidade de dados do mundo humano—da psicologia, em particular—pode gerar estranheza.

O psicologismo, a tese de que os fazedores de verdade das leis lógicas sejam fatos sobre entidades mentais, é considerado "morto e enterrado", por assim dizer, desde Frege (1964; 1979). Não pretendo ressuscitar o psicologismo aqui, até porque a evidência empírica disponível mostra que o psicologismo estava, sim, errado, como veremos. Mas, mesmo que fatos sobre a mente não sejam os fazedores de verdade das leis lógicas, ainda assim dados da psicologia são relevantes para testar outras hipóteses sobre os fazedores de verdade da lógica. Considere, por exemplo, três das hipóteses mais comumente discutidas sobre os fazedores de verdade das leis lógicas:

a) hipótese linguística, segundo a qual a lógica captura aspectos estruturais da linguagem;
b) a hipótese racionalista, segundo a qual a lógica captura os modos de pensar que caracterizam a racionalidade humana; e
c) a hipótese empirista, segundo a qual a lógica captura aspectos estruturais do mundo.

Estas três hipóteses têm consequências empíricas testáveis, como mostrarei no restante desta seção. Para as três, resultados da psicologia são importantes, como veremos nas seções 4, 5 e 6. Como há uma diversidade considerável de especificações dos detalhes de cada uma destas hipóteses, dependendo do autor que a defende, vou eleger um representante de cada uma delas: Carnap, para a hipótese linguística; Béziau, para a hipótese racionalista; e Williamson, para a hipótese empirista. Da proposta de cada um desses autores, derivarei uma consequência empiricamente testável.

Carnap e a hipótese linguística

Na versão carnapiana, a hipótese linguística explica as verdades lógicas por meio da ideia de analiticidade. Uma sentença é analítica se sua verdade depende somente do significado dos termos nela empregados. Por exemplo, a sentença 'Maria é casada ou Maria é solteira' pode ser entendida como analiticamente verdadeira se por 'solteira' queremos dizer tão somente 'não-casada'. Neste caso, a sentença é verdadeira unicamente em função do significado das palavras 'ou', 'é' e 'solteira', independentemente de quem seja a Maria da qual estamos falando. Agora tome a sentença 'Todas as pessoas são casadas ou são não-casadas'. Esta sentença pode ser vista como uma instância do princípio do terceiro excluído e formalizada como $\forall x(\mathrm{C}x \vee \neg \mathrm{C}x)$, onde C é a propriedade de ser casada. Repare que esta fórmula (na sua interpretação clássica padrão) será verdadeira independentemente de como interpretamos a propriedade C. De acordo com a hipótese linguística, esta sentença é analítica e verdadeira unicamente em função dos termos *lógicos* nela empregados: quantificador universal, disjunção e negação. Poderíamos estar falando de qualquer outra coisa, não relacionada com estado civil, e a mesma forma lógica produziria sentenças verdadeiras.

Até aqui, tudo bem, se aceitamos o conceito de analiticidade (Quine (1976) tem críticas duras a ele). Há, contudo, um segundo aspecto da versão carnapiana da hipótese linguística, que é sua explicação da origem dos significados dos termos lógicos. De acordo com Carnap, o significado dos termos lógicos é estabelecido por meio de convenções; em outras palavras, escolhemos

livremente que lógica vamos usar. Esta lógica é introduzida por meio de uma estrutura linguística que estipula os termos da linguagem e suas regras de uso. Estas estipulações, segundo Carnap, não respondem a nenhuma realidade subjacente, seja ela física ou mental; elas seriam feitas em bases unicamente pragmáticas. Uma vez estabelecidas, porém, as regras da estrutura linguística têm certas implicações, de forma que algumas construções dentro da estrutura linguística resultam verdadeiras unicamente em função das regras que regem os termos nelas empregados; isto é, tornam-se verdades analíticas. Para Carnap, este seria o caso das verdades lógicas e matemáticas (Carnap, 1983; Hylton, 2007, p. 43 e seguintes).

Retornando ao exemplo acima—$\forall x(\mathrm{C}x \vee \neg \mathrm{C}x)$—, repare que, em sua interpretação padrão, o quantificador universal desta sentença quantifica sobre um domínio de objetos, com propriedades e em relações entre si. Sendo assim, os conceitos de objeto e propriedade contribuem tanto para a validade (verdade em todos os modelos) desta fórmula quanto os conceitos de quantificação universal, disjunção e negação. Então, para sermos mais específicos, precisamos dizer que a verdade analítica de '$\forall x(\mathrm{C}x \vee \neg \mathrm{C}x)$' deriva não só dos conceitos de quantificação universal, disjunção e negação, mas também dos conceitos de objeto e propriedade. Por exemplo, esta sentença resulta verdadeira analiticamente somente se é uma regra da estrutura linguística que, para todo par composto por um objeto x e uma propriedade P, jamais fica indefinido se x é ou não é P. Desta forma, em uma estrutura linguística que valide '$\forall x(\mathrm{C}x \vee \neg \mathrm{C}x)$' os conceitos de objeto e propriedade têm de ser introduzidos com as regras adequadas, ou esta sentença não será analiticamente verdadeira. De fato, em Carnap (1983), Carnap dá como exemplo de uma estrutura linguística a "linguagem das coisas", que introduz substantivos e suas regras de uso, e o "sistema das propriedades das coisas", um acréscimo à linguagem das coisas que introduz adjetivos e suas regras de uso.

Quer dizer, para Carnap, tanto os conceitos lógicos propriamente ditos, quanto os conceitos de objeto e propriedade, essenciais para lógicas que quantificam sobre objetos, são introduzidos linguisticamente. Agora, pedindo perdão aos carnapianos pelas simplificações, estamos na altura certa para deri-

var da hipótese linguística carnapiana uma consequência empírica. Se estes conceitos são todos linguísticos, quem não adotou uma estrutura linguística que os introduza, não deveria tê-los. Se encontrarmos seres que não falam uma linguagem desse tipo, mas que ainda assim exibam conhecimento, por exemplo, dos conceitos de objeto e propriedade, a hipótese carnapiana estará refutada. Bebês humanos em sua fase pré-linguística são seres desse tipo. Na seção 4 veremos como bebês humanos demonstram conhecimento dos conceitos de objeto e propriedade.

Antes de prosseguirmos, um aviso: é evidente que há inúmeras maneiras legítimas de resistir a esta possível refutação; por exemplo, pode-se argumentar que Carnap tinha em mente linguagens artificias, e não linguagens naturais; pode-se argumentar que já nascemos com uma linguagem, a "linguagem do pensamento" (Fodor, 1975); ou pode-se dizer que investigar conceitos pré-linguísticos é uma pseudo-questão. Como meu propósito aqui é apenas ilustrar um método empiricamente informado de abordar problemas em filosofia da lógica, e não resolvê-los, não abordarei essas questões.

Béziau e a hipótese racionalista

Em Béziau (2010), Béziau observa que o termo 'lógica' é ambíguo. Segundo ele, devemos distinguir entre "lógica como raciocínio", que ele grafa com inicial maiúscula—Lógica—e "lógica como o estudo do raciocínio", isto é, a disciplina desenvolvida pelos lógicos, que ele grafa com inicial minúscula. E ele arremata: "*Lógica* [é] o objeto de estudo da *lógica*" (Béziau, 2010, p. 76). Em outras palavras: o raciocínio é o objeto de estudo da lógica. Temos aí a hipótese racionalista: a lógica não é sobre a linguagem, nem sobre o mundo, mas sobre uma atividade humana, notadamente, o ato de raciocinar. É importante notar como a posição de Béziau se distancia tanto do psicologismo, que veria o ato de racionar como um processo mental, privado e subjetivo, quanto de uma forma de platonismo que veria a razão como algo de outro mundo. Isso fica claro na analogia que Béziau traça entre lógica e dança.

> Dança é uma arte e uma prática desta arte […] podemos dizer: "Mary está estudando dança para dançar". E quanto a John? Diríamos: "John está estudando lógica para raciocinar". […] Na […] dança e na lógica, há por um lado uma atividade desempenhada por seres humanos […] e, por outro, uma teoria sobre esta atividade produzida pelos mesmos mamíferos (Béziau, 2010, p. 77, minha tradução).

O objeto de estudo da lógica é uma *atividade* humana, desempenhada externamente, em espaço público, tal qual o objeto de estudo da dança (como teoria) é uma atividade humana, a dança. Dançamos num palco ou num salão, e raciocinamos em diálogos, discursos, ou privadamente; mas, mesmo que privadamente, o produto da prática racional é externamente acessível: um texto, uma teoria científica ou matemática, um discurso elaborado.

Estas atividades racionais, que constituem o que Béziau chama de Lógica, são o objeto de estudo da lógica, a disciplina dos lógicos. Isto está de acordo com o princípio de sistematização de da Costa, que mencionei acima, e está na base da ideia de que a lógica não é uma ciência excepcional, isto é, que a lógica não ocupa uma posição *a priori* com respeito às demais ciências. Ao contrário, na concepção de da Costa a lógica é uma disciplina científica *a posteriori* como qualquer outra (da Costa & Arenhart, 2018). Ela se ocupa de investigar um fenômeno no mundo, qual seja, a razão humana. Em linha com da Costa, na sequência da passagem citada acima, Béziau se pergunta: "mas é a lógica uma ciência humana?" (Béziau, 2010, p. 77, minha tradução).

Aristóteles define o ser humano como o animal racional. "Esta definição implica que a racionalidade é um traço essencial dos seres humanos e que, consequentemente, os humanos sempre foram racionais" (Béziau, 2010, p. 81, minha tradução). Como Béziau nota, 'razão' é a tradução latina para o grego 'logos' e, portanto, uma tradução literal da definição aristotélica de humano seria "animal lógico".

Agora, pedindo perdão a Béziau pelas simplificações, chegou o momento de derivarmos uma consequência empírica de sua hipótese racionalista. Se o ser humano é o "animal lógico", então podemos supor que, quando humanos raciocinam corretamente, eles os façam de um modo compatível com os cânones lógicos. Portanto, pelo menos alguns dos sistemas lógicos propos-

tos pelos lógicos deveriam ser capazes de modelar adequadamente os modos de pensar que caracterizam a racionalidade humana; afinal, este é o objeto de estudo da lógica, segundo Béziau. Esta é uma questão que pode ser estudada empiricamente: se observarmos atenta e sistematicamente como os humanos raciocinam na prática, podemos avaliar se eles de fato seguem cânones lógico-dedutivos ou não. Como veremos na seção 6, os psicólogos já fizeram isso, e a resposta parece ser desfavorável para a hipótese racionalista. (Novamente, há muitas maneiras legítimas de resgatar a hipótese racionalista diante desta possível contra-evidência, mas não as abordarei aqui.)

Williamson e a hipótese empirista

Ao passo que para Carnap e Béziau o objeto de estudo da lógica é um fenômeno bem determinado—estruturas linguísticas para o primeiro, a atividade de raciocinar para o segundo—, para Williamson (2013; 2017) o objeto de estudo da lógica compreende absolutamente tudo. G. Russell (2018) chama a concepção de Williamson de "universalista", no sentido de que, na visão de Williamson, uma sentença é uma verdade lógica se e somente se ela for verdadeira sobre toda e qualquer coisa que existe no universo. Por exemplo, o princípio do terceiro excluído é uma verdade lógica, de acordo com Williamson, porque dados qualquer propriedade X e qualquer objeto x, é o caso que x é X ou x não é X (isto é, $\forall X \forall x(Xx \vee \neg Xx)$). Para que o princípio do terceiro excluído seja uma verdade lógica, isto tem de valer tanto para objetos matemáticos quanto para objetos ficcionais, objetos físicos, e qualquer outro tipo de objeto existente. Nesta concepção, a lógica é a ciência que estuda as propriedades mais gerais do universo; ao passo que a física estuda as propriedades do mundo físico, e a matemática estuda as propriedades de certos objetos abstratos, a lógica estuda as propriedades que há em comum entre estes e todos os demais tipos de objetos. É nesse sentido que podemos dizer que, para Williamson, a lógica captura aspectos estruturais do universo, concebido de maneira ampla. O universo assim concebido engloba não só o mundo físico, mas *também* o mundo físico, de sorte que as leis lógicas valem, nesta concepção, apenas se elas são verificadas, em particular, pelos objetos físicos e suas

propriedades. Neste sentido, toda lei lógica tem uma dimensão empírica: ela pretende ser verdadeira sobre o mundo físico e, portanto, pode ser testada contra a observação empírica. Na seção 5, veremos como a relação entre leis lógicas e o mundo físico pode ser mobilizada contra a hipótese empirista.

4. Testando a hipótese linguística contra os dados

Como vimos, a introdução de conceitos lógicos e dos conceitos de objeto e propriedade, segundo Carnap, se dá por meio da adoção de uma estrutura linguística. Como sugeri, isto tem uma consequência empírica testável: quem não adotou uma estrutura linguística que inclua esses conceitos não deveria conhecê-los.

Na psicologia do desenvolvimento, há duas hipóteses principais para explicar a origem ontogênica (isto é, na história de desenvolvimento de cada indivíduo) do conceito de objeto: a hipótese nativista e a hipótese do aprendizado. Nenhuma delas atribui qualquer papel à linguagem na aquisição desses conceitos. Todos os resultados que reporto no restante desta seção vêm de Johnson (2013) e Johnson (2020), exceto quando outra fonte é citada.

Note que aqui estou entendendo que conceitos são entidades psicológicas, no sentido definido por Machery (2017), mencionado acima. Um conceito psicológico não precisa estar formulado em uma linguagem, e por isso existe a possibilidade de alguém ter o conceito de objeto ainda que não fale a linguagem das coisas ou outra qualquer. "A maneira correta de articular o conteúdo de um conceito é pedir para as pessoas usarem-no (por exemplo, pedir para aplicá-lo) e inferir seu conteúdo do seu uso" (Machery, 2017, p. 210). Assim, mesmo agentes que não falem linguagem alguma podem ter seus conceitos investigados, desde que se encontre maneiras criativas de solicitá-los a empregar o conceito em questão, caso o tenham. Isto é o que vemos em estudos com bebês: os pesquisadores criam situações em que os bebês são solicitados a recrutar seu conceito de objeto, caso o tenham. E os resultados mostram que, de fato, eles têm uma compreensão bastante significativa sobre objetos, indicando assim que o conceito de objeto é pre-linguístico, como veremos a seguir. Talvez, para Carnap, conceitos sejam entidades exclusivamente linguís-

ticas, não-psicológicas, e então um resultado como este não seja uma "refutação" de sua hipótese; ou talvez para Carnap perguntar sobre a natureza dos conceitos seja uma pseudo-questão. Não abordarei esses pontos aqui; para extrair uma consequência empírica de uma teoria que não pretendia tê-las, é preciso fazer adaptações. O importante é estar claro que adaptações estão presentes. Passemos aos resultados científicos.

Segundo a hipótese nativista, o conceito de objeto é inato, isto é, já faz parte da configuração básica do aparelho cognitivo humano, por assim dizer. Há vários resultados experimentais que parecem confirmar a hipótese nativista. Os psicólogos costumam decompor o conceito de objeto em quatro dimensões: unidade; persistência; coerência; e identidade. Completo domínio do conceito de objeto envolve demonstrar habilidades correspondentes a cada uma dessas dimensões. Unidade é testada solicitando que o sujeito sob teste demonstre habilidade de reconhecer que as pontas de um objeto cujo centro foi parcialmente ocultado constituem um único objeto. Persistência, por meio da habilidade de reconhecer que um objeto totalmente ocultado continua a existir. Coerência, por meio da habilidade de estranhar figuras de objetos impossíveis, como o triângulo de Penrose ou as escadas de Escher. Identidade, por meio da capacidade de reconhecer um objeto encontrado anteriormente como sendo o mesmo num encontro posterior. Surpreendentemente, bebês de apenas quatro meses, portanto muito antes de balbuciarem as primeiras palavras, demonstram todas essas habilidades em algum grau.

Leitores não familiarizados com estudos da psicologia do desenvolvimento podem estar se perguntando como os pesquisadores sabem disso. A maior parte desses estudos usa um método de experimentação com bebês chamado "violação de expectativa": o bebê é primeiramente habituado com um certo estímulo, e o tempo que ele olha para este estímulo é medido; em seguida, ele é apresentado a um novo estímulo. Se o bebê olha para esse novo estímulo por mais tempo, isto é interpretado como um renovado interesse, indicando que o bebê notou uma diferença. Por exemplo, para testar a percepção de coerência, o bebê pode ser habituado com a figura de um objeto possível e, em seguida, exposto a uma versão impossível do mesmo objeto. Olhar

mais longamente para o objeto impossível, depois da habituação com o objeto possível, demonstra que ele notou a diferença.

Mas nem todas as habilidades de um adulto estão presentes em bebês, sugerindo que algo tem de ser aprendido. Os resultados experimentais que suportam a hipótese do aprendizado mostram que, embora os bebês demonstrem as habilidades mencionadas acima por volta dos quatro meses de idade, essas habilidades ainda não se encontram no nível observado em adultos ou crianças mais velhas. A percepção de identidade ilustra bem este ponto. Bebês de três meses e meio baseiam-se quase que somente na velocidade e caminho percorrido por objetos para identificar a presença de objetos distintos. Por volta dos quatro meses eles começam a usar outras propriedades, tais como tamanho, forma e som, mas só muito mais tarde vão usar cor e brilho para identificação. Isso pode indicar um necessário período de maturação de estruturas cognitivas inatas, mas também pode indicar a necessidade de aprendizado. Há vários outros resultados, principalmente ligados ao aprendizado advindo da manipulação ativa de objetos em um espaço tridimensional, que corroboram a necessidade de aprendizado. Por exemplo, há estudos que mostram que bebês que são exploradores manuais mais ativos tendem a reagir com estranheza (maior tempo de olhar) a estímulos onde objetos desconectados movem-se juntos.

Como se nota, esses mesmos resultados ilustram a capacidade dos bebês de perceber propriedades de objetos (velocidade, forma, etc.), usá-las para reidentificá-los, e de perceber relações entre objetos (conectados/desconectados). Tudo isso antes de adquirir linguagem. Isso mostra que os conceitos de objeto, propriedade e relação não têm origem linguística. Na verdade, pode-se dizer que falamos sobre eles na linguagem *porque* já os tínhamos antes de sabermos falar. Mais que isso, é bem possível que esses conceitos já andassem por aí muito antes do surgimento do *homo sapiens*, pois primatas não-humanos têm mostrado habilidades bastante desenvolvidas com respeito aos conceito de objeto e propriedade (Mendes, Rakoczy & Call, 2008).

Em face da evidência empírica, é forçoso admitir que os conceitos de objeto e propriedade são pré-linguísticos. Pode até ser que uma sentença lógi-

ca tal qual '$\forall x(Cx \vee \neg Cx)$' seja válida unicamente em função do significado dos seus termos, mas alguns destes significados não têm origem na linguagem, diferente do que propunha Carnap. A introdução da linguagem das coisas e do sistema de propriedades parece antes responder a fatores pré- e extra-linguísticos: é porque nosso sistema perceptivo lida com o mundo usando essas categorias que introduzimos na linguagem termos gerais como 'objeto' e 'propriedade' para designar estes tipos de entidades.

Note que a contra-evidência apresentada aqui diz respeito à linguagem das coisas apenas, em especial aos conceitos de objeto e propriedade. Na medida em que essas noções são empregadas em certos sistemas lógicos, e dada sua origem pré-linguística, podemos dizer que tais sistemas lógicos não são exclusivamente sobre conceitos linguísticos. Mas isso nada diz sobre os conceitos lógicos propriamente ditos, como a noção de implicação. Neste ponto, Carnap parece ter razão. Como veremos na seção 6, estes conceitos, no uso especializado que se faz deles em lógicas dedutivas, precisam ser explicitamente introduzidos antes que alguém tenha pleno domínio deles.

Fica a pergunta sobre a origem das conceitos pré-linguísticos de objeto e propriedade usados pelo nosso sistema perceptivo. Alguém poderia dizer que a hipótese nativista acaba vindicando pelo menos um aspecto da hipótese carnapiana. Embora a origem destes conceitos não esteja numa estrutura linguística criada por humanos, ela poderia estar numa espécie de linguagem do pensamento, sendo, portanto, *a priori*. Mas essa conclusão não pode ser tirada tão rapidamente. Primeiro, porque esses conceitos não são totalmente inatos, visto que o aprendizado cumpre ainda um papel no desenvolvimento dos conceitos adultos, como vimos. A compreensão atual é que a hipótese nativista e a hipótese do aprendizado são complementares. Mas mesmo o componente inato do conceito de objeto não parece ser *a priori*. É bem provável que ele tenha uma origem evolutiva, ligado a pressões seletivas do ambiente que favoreceram indivíduos para os quais a capacidade de lidar com objetos constituiu uma vantagem. Desta forma, mesmo o conteúdo inato teria origem na experiência; não na experiência do indivíduo que o porta, mas na experiência de uma linhagem de antepassados (Knudsen, 2020).

5. Testando a hipótese empirista contra os dados

O último parágrafo da seção anterior sugere que os conceitos de objeto, propriedade e relação possam ter origem externa, no mundo: seria porque há objetos lá fora que seres com a capacidade de perceber objetos teriam uma vantagem evolutiva. Esta hipótese parece tentadora, e é certamente plausível a partir de uma perspectiva científica. Mas ela é filosoficamente questionável. Lembre-se do direito que o filósofo empiricamente informado se reserva de criticar a ciência, sem para isso precisar recorrer à filosofia primeira. Como veremos, há resultados científicos que podem ser usados para reforçar a dúvida sobre quanto do modo como organizamos nossa percepção provém do mundo exterior.

O que se sabe de mais seguro sobre como o cérebro forma imagens é que existe uma grande quantidade de processamento e transformação entre os estímulos recebidos na retina e as imagens que vemos. Para começo de conversa, o estímulo recebido na retina é bidimensional, o que significa que nossa visão de objetos tridimensionais é resultado de uma transformação operada pelo cérebro (Pylyshyn, 1999, p. 354). Para fazer esse processamento, o cérebro utiliza informações que não estão contidas na imagem retinal (Teufel & Nanay, 2016). A partir daí, há pouco consenso entre os cientistas que investigam a percepção visual sobre a fidelidade das imagens geradas com respeito ao mundo exterior. *Grosso modo*, há duas correntes antagônicas. Para uns, as informações que o cérebro usa para processar os estímulos recebidos pela retina apenas restauram dados reais sobre o mundo que, devido a limitações da retina, não são capturados como dados brutos de entrada. Para outros, nem sempre é isso que ocorre, havendo uma influência significativa de fatores não-perceptuais, tais como crenças e expectativas, no processamento das imagens retinais.

Teufel & Nanay (2016) descrevem em linhas gerais como funcionaria a percepção visual de acordo com os modelos da primeira corrente.

> [As] computações [operadas pelo cérebro sobre a imagem retinal] dependem significativamente de informações que não estão contidas na imagem retinal. Os sinais que emanam do ambiente são estruturados e apresentam certas regularidades […] Muitos organismos ex-

> ploram este fato ajustando com precisão seus sistemas sensoriais e perceptuais à estrutura das características do ambiente que são relevantes para a sua sobrevivência e reprodução. Para citar dois exemplos, as propriedades de campo receptivo linear e não linear dos neurônios no sistema visual inicial dos primatas estão estruturadas de forma a codificar o ambiente natural da forma mais eficiente possível […], e os processos de integração de contornos em humanos refletem as estrutures de contorno de objetos em cenários naturais […] Estes exemplos mostram que as estrutura dos sistemas biológicos de visão refletem as propriedades relevantes do ambiente em que os organismos vivem. Em outras palavras, a informação prévia sobre as regularidades estatísticas no ambiente de um organismo é embutida na forma como o sistema visual processa os sinais recebidos (Teufel & Nanay, 2016, p. 5, minha tradução).

Perceba que, neste modelo, o cérebro suplementa a informação que não entra pela retina, mas isso é feito por meio de informações previamente codificadas no sistema visual que corresponderiam a dados do mundo exterior. Neste caso, leis lógicas que nos permitem fazer inferências sobre objetos, suas propriedades e relações poderiam ter um pé na realidade proporcional a quanto do conceito de objeto (isto é, a ideia de que o mundo é segmentado o em objetos) advém do mundo exterior.

Para os cientistas ligados à segunda corrente, porém, fatores originados de funções cognitivas superiores, e mesmo fatores culturais podem penetrar o processamento visual e mudar o que vemos (Lupyan, Thompson-Schill & Swingley, 2010; Vetter & Newen, 2014). Certamente, há limites para estas modificações impostas "de cima para baixo", mas ainda assim as modificações podem ser relevantes. Por exemplo, objetos desejados pelo observador ou considerados ameaçadores podem ser percebidos como fisicamente mais próximos, ao passo que objetos menos desejados ou repugnantes são percebidos como mais distantes (Vetter & Newen, 2014). Segundo Vetter & Newen (2014), efeitos como esse acontecem porque, no processo de formar a imagem que vemos, o cérebro primeiro cria uma imagem preliminar, que é confrontada com esquemas visuais aprendidos, e portanto expostos a influências culturais. Se a imagem preliminar não "encaixa" no esquema visual aprendido, ela é modificada, e então um novo encaixe com esquemas visuais aprendidos é tentado. Quando esse processo estabiliza, temos a imagem final, resultante tanto dos estímulos externos como de fatores internos devidos a funções cog-

nitivas superiores. Para ter um exemplo prático disso, veja a Fig. 1 antes de ler a citação abaixo.

> [A]o ver a imagem pela primeira vez, a maioria das pessoas percebe apenas um padrão não estruturado de manchas pretas num fundo branco. Depois de um certo tempo ou depois de ver uma imagem em resolução mais alta, todos percebem o dálmata mesmo na imagem em baixa resolução. Aqui, o modelo memorizado de um cão modifica a percepção visual de um padrão de ruído aparentemente não estruturado (Vetter & Newen, 2014, p. 69, minha tradução)

Quanto da segmentação do campo visual em objetos é similar à percepção do dálmata na Fig. 1? Se esse efeito for relevante e perpassar a maior parte da nossa experiência visual, é possível argumentar que a organização do mundo em objetos com propriedades e em relações entre si é mais uma imposição da nossa mente que uma característica do mundo. Neste sentido, leis lógicas calcadas nos conceitos de objeto e propriedade não capturariam aspectos estruturais do mundo. Elas descreveriam mais nosso modo de ver o mundo do que o mundo propriamente dito. E isso vai contra a hipótese de Williamson, de acordo com a qual as leis lógicas são verdadeiras (se verdadeiras) sobre o próprio mundo físico, e não sobre nossa percepção do mundo físico.

Fig. 1. Manchas pretas num fundo branco que, alguns segundos depois, "encaixam" no padrão de um cão dálmata(no centro, de costas, com o focinho abaixado em direção ao solo). (Imagem da web.)

Em consonância com a hipótese de que há um descompasso entre como vemos o mundo e o mundo propriamente dito, há evidências vindas da física que parecem indicar que, em última instância, o mundo de fato não se estrutura em objetos. É bem sabido que o mundo quântico não se comporta como o mundo macroscópico. Partículas sub-atômicas não exibem o comportamento dos objetos com os quais estamos acostumados. Não preciso me demorar neste ponto porque o leitor provavelmente já está familiarizado com o influente trabalho de Newton da Costa, Décio Krause e outros neste tópico (veja, p. ex., French & Krause, 2003; da Costa & Krause, 1994; e da Costa, 1994). Mas, só para refrescar a memória, vale lembrar do experimento da dupla fenda, no qual elétrons são disparados contra um anteparo que fica marcado onde um elétron o atinge. Se o fluxo de elétrons é feito passar por uma barreira com uma única fenda, o padrão registrado no anteparo pelos elétrons que o atingem é consistente com o comportamento de partículas—isto é, bolinhas de gude deixariam um padrão semelhante. Mas, se a barreira possuir duas fendas, o modo como os elétrons atingem o anteparo é consistente com o comportamento de ondas—isso jamais aconteceria com bolinhas de gude. Este resultado, somado a outras coisas estranhas do mundo quântico, sugere que nossa categoria "objeto" não se aplica no âmbito sub-atômico. Baseada neste resultado, Maddy conclui:

> No final, parece que devemos abrir mão da convicção de que cada elétron viaja através de uma fenda ou da outra; devemos abrir mão da convicção de que um elétron tem, a cada momento, uma posição no espaço; devemos abrir mão da convicção de que um elétron, ao deixar sua origem em um momento, e atingir o anteparo um curto tempo depois, tenha percorrido um caminho contínuo de um ponto a outro (Maddy, 2007, p. 236, minha tradução).

Isto é, devemos abrir mão de que elétrons sejam objetos e se comportem como tais. Por conseguinte, pelo menos no nível sub-atômico, sabemos que nosso mundo não é estruturado em objetos (ao menos esta é a posição que Maddy(2007) defende). Mais uma razão para se dizer que leis lógicas que dependem da noção de objeto não refletem o mundo *simpliciter*. Mas, fazendo as devidas qualificações, ainda podemos afirmar, como fiz acima, que leis lógicas refletem aspectos estruturantes da nossa *percepção* do mundo *macroscópico*.

Quase isso; mais uma qualificação ainda é necessária. Afinal, mesmo no nível macroscópico temos razões para dizer que existem muitas coisas que não são objetos. Um objeto típico, aquele cuja percepção é estuda pela psicologia, é um corpo com bordas bem definidas, exibindo unidade, persistência, coerência e identidade, como mencionei acima, seguindo Johnson (2013). *Eventos*, por exemplo, não são objetos: eles não têm bordas, eles são transitórios, em vez de persistentes, e cada evento é único, não reencontramos um evento novamente como reencontramos objetos (o máximo que reencontramos são eventos do mesmo tipo). Ademais, eventos naturais, como as ondas do mar ou uma rajada de vento, não parecem ter bordas definidas. Logo, é forçoso admitir que pelo menos uma parte da realidade não é estruturada em objetos.

Mais ainda, se olhada bem de perto, mesmo a parte que nos aparece como estruturada em objetos não encaixa perfeitamente no esquema representado no domínio de uma estrutura que interpreta uma linguagem de primeira ordem. Nestes domínios, para cada objeto está determinado se ele possui ou não uma certa propriedade ou relação. Não há espaço para indeterminação. Acontece que muitas vezes o mundo físico é indeterminado quanto às propriedades e relações de seus objeto. Maddy (2007, p. 240 e seguintes) usa esse fato para rejeitar que leis como o terceiro excluído sejam verdadeiras sobre o mundo físico em geral. Os exemplos se multiplicam: não é possível determinar, exceto por critérios em certa medida arbitrários, o limite entre alguém ser careca ou não (exatamente quantos fios de cabelo alguém precisa ter para não contar como careca?); o momento exato em que um girino se torna um sapo; o ponto exato em que uma montanha termina e começa outra em uma cadeia de montanhas; o ponto exato onde passa a borda de um vale, de sorte que pode ser indeterminado se uma certa árvore tem a propriedade de pertencer ao vale ou não; e assim por diante. Sendo esse um fenômeno bem comum no nosso mundo, parece inadequado dizer que leis lógicas que dependem de que toda propriedade seja bem determinada com respeito a todos os objetos, reflitam a estrutura do mundo. No máximo, pode-se dizer que uma lei tal qual o princípio do terceiro excluído é verdadeira, sem requerer idealizações e arbi-

trariedades, somente sobre nossa experiência com setores específicos do mundo, notadamente aqueles onde estamos lidando com objetos genuínos sem vaguidade (os objetos macroscópicos da minha mesa, por exemplo, parecem satisfazer essas condições, se restringimos o universo ao que é percebido sobre a minha mesa; os acidentes geográficos do Planalto Central, não).

A resposta padrão para aqueles que, como Williamson, defendem a verdade irrestrita (i.e., em todos os caso) de uma sentença como '$\forall X \forall x(Xx \vee \neg Xx)$' é que a vaguidade é um problema epistêmico: nós humanos podemos ser incapazes de determinar a linha exata onde passa a borda de um vale, mas este limite existe (Williamson, 1994). Há um enorme debate filosófico sobre a vaguidade, no qual não cabe entrar aqui (veja, p. ex., Sorensen (2018)). De toda forma, as considerações feitas acima sobre as influência de funções cognitivas superiores no nosso sistema perceptivo sugerem que as categorias de objeto e propriedade possam ter um componente devido ao nosso sistema cognitivo, o que torna menos plausível a ideia de que há uma linha exata determinando as bordas de um vale independentemente do que achemos sobre isso. A própria segmentação da percepção em vale e montanha pode já ser, ao menos em parte, uma imposição nossa ao ambiente.

Recapitulemos o que vimos até aqui. Os conceitos de objeto e propriedade não têm origem linguística (embora conceitos mais propriamente lógicos, como a disjunção e a negação, possam ter). É difícil afirmar taxativamente que estes conceitos reflitam aspectos universais estruturantes do mundo, mas parece muito menos controverso afirmar que eles refletem aspectos estruturantes do modo como percebemos setores específicos da realidade macroscópica. Por conseguinte, leis lógicas que articulam esses conceitos podem estar capturando tais aspectos estruturantes da nossa relação com esses setores da realidade.[2] Sendo assim, quem organiza o mundo em objetos, propriedades e relações bem definidos, estáveis e bem comportados como aqueles da lógica clássica somos nós. Diante disso, a hipótese empirista perde força.

[2] Maddy (2007) defende que há um subconjunto da lógica clássica, o qual ela chama de lógica rudimentar, que vale em geral no mundo tal qual percebido por nós.

6. Testando a hipótese racionalista contra os dados

É um fato, como se depreende das seções anteriores, que pensar em termos de objetos com propriedades e relações entre si é algo bastante intuitivo para nós, algo que estrutra a percepção e a cognição desde o nascimento, praticamente. Contudo, por certo isso não é tudo de que se precisa para se obter uma lógica. Leis lógicas estabelecem relações lógicas entre objetos ou proposições, as quais nos permitem fazer inferências dedutivas sobre eles. A hipótese racionalista, como vimos, diz que, quando humanos—o "animal lógico", na expressão de Béziau (2010)—raciocinam corretamente sobre os objetos e proposição que eles encontram no dia a dia, eles o fazem de modo compatível com as regras especificadas por sistemas lógico-dedutivos. Esses sistema lógicos apenas codificariam essas maneiras de raciocinar já praticadas comumente pelos humanos. Se for este o caso, deveria ser fácil encontrar inferências dedutivas logicamente válidas entre os raciocínios que pessoas comuns, não treinadas em lógica, fazem no seu dia a dia.

Em meados do século XX, psicólogos começaram a investigar o pensamento humano tomando como modelo de racionalidade a lógica clássica. De certa forma, é como se eles tivessem adotado a hipótese racionalista, e assumido que a lógica clássica fosse o sistema lógico que capturasse o modo como pensamos corretamente. Um dos métodos de teste mais comuns empregado pelos psicólogos consistia em apresentar silogismos e solicitar aos participantes do estudo que avaliassem se a conclusão seguia-se logicamente das premissas. Como os professores costumam instruir seus alunos em cursos introdutórios de lógica, os participantes desses estudos eram explicitamente instruídos pelos experimentadores a assumir que as premissas eram verdadeiras, mesmo que porventura fossem contrárias a suas crenças ou flagrantemente falsas, e avaliar a conclusão com respeito unicamente àquelas premissas. Em um estudo desse tipo, Sá, West & Stanovich (1999) apresentaram aos participantes—pessoas que receberam uma boa formação escolar mas que nunca frequentaram um curso de lógica—o seguinte silogismo:

Todas as coisas vivas precisam de água;
Rosas precisam de água;
Portanto, rosas são coisas vivas.

Solicitados a avaliá-lo, 68% dos participantes julgaram este silogismo como válido, embora qualquer pessoa treinada em lógica o reconheça facilmente como inválido (um caso da falácia da afirmação do consequente). O que este resultado constata é aquilo que todo professor de lógica percebe nos primeiros dias de aula de uma nova turma: silogismos cuja conclusão é plausível tendem a ser tomados como válidos (mesmo que a conclusão não siga logicamente das premissas), ao passo que silogismos cuja conclusão é implausível tendem a ser tomados como inválidos (ainda que a conclusão se siga logicamente das premissas). Este resultado é confirmado em outros estudos do mesmo tipo. Por exemplo, em Evans, Barston & Pollard (1983), 71% das avaliações de argumentos inválidos com conclusão plausível foram positivas (foram julgados válidos), enquanto apenas 56% das avaliações de argumentos válidos mas com conclusão implausível foram positivas (quer dizer, em 44% das vezes, os participantes falharam em reconhecer um argumento válido porque ele tinha uma conclusão implausível).

Os psicólogos têm chamado esse efeito de *viés de crença* (*belief bias*, em inglês): "conclusões fáceis de acreditar são consideradas mais aceitáveis que conclusões difíceis de acreditar, independentemente da validade lógica" dos argumentos usados para suportá-las (Ball & Thompson, 2018, p. 16, minha tradução). A ideia é que o agente usa suas crenças para avaliar a plausibilidade da conclusão; se a conclusão está de acordo com suas crenças, o agente tende a considerar o argumento válido. É fácil ver que o viés de crença fere a monotonicidade característica das inferências dedutivas. O agente traz informações que não estão nas premissas para avaliar o argumento, de forma que a classificação do argumento como válido ou inválido pode variar conforme as crenças particulares de cada indivíduo.

O viés de crença é intimamente relacionado ao chamado *viés de confirmação* (*confirmation bias*) ou *viés do meu-lado* (*myside bias*): "a tendência

[que as pessoas têm] de encontrar argumentos que suportem seus pontos de vista, seja para apoiar uma posição com a qual elas concordam ou para atacar uma posição da qual elas discordam" (Mercier, 2018, p. 404, minha tradução). Quer dizer, assim como tendemos a avaliar positivamente argumentos cuja conclusão concorda com nossas crenças, temos também mais facilidade em encontrar argumentos para defender nossas crenças do que argumentos contra elas. Assim, ao considerar um argumento diante do qual temos que ativamente procurar por contra-exemplos (para testar sua validade), enfrentamos maior dificuldade em fazê-lo se concordamos com sua conclusão. Daí que argumentos inválidos com conclusões aceitáveis passem despercebidos.

A consistência desses resultados, confirmados em vários estudos (para uma revisão da literatura, veja Ball & Thompson, 2018), mostra que o nosso modo espontâneo de raciocinar diverge do que seria esperado logicamente.[3] Isto reforça a hipótese de que o modo de pensar dedutivo não é nosso modo "padrão" de raciocinar. Pessoas não treinadas em lógica tendem a não usar raciocínio dedutivo mesmo quando confrontadas com um raciocínio dedutivo (um silogismo, nos experimentos mencionados acima). A conclusão que se segue é que precisamos *aprender* a fazer inferências dedutivas, e isso exige consideráveis horas de treinamento (um curso introdutório de lógica, ao menos).

Um outro tipo de estudo muito citado para mostrar como nosso modo espontâneo de pensar diverge do esperado logicamente é a chamada tarefa de seleção de Wason. Esta tarefa visa testar a compreensão da implicação material. Conforme descrito em Elqayam (2018), aos participantes testados nesta tarefa são apresentadas quatro cartas, sendo que cada carta tem uma letra escrita de um lado e um número, do outro. As cartas são dispostas sobre a mesa, deixando apenas um dos lados visível. Por exemplo, os participantes veem diante de si as cartas N, T, 6, 8. Em seguida, os participantes são solicitados a verificar se a seguinte condicional é verdadeira sobre as cartas que têm diante de si: "Se uma carta tem um N de um lado, então ela tem um 6 do ou-

[3] Se o parâmetro adotado for uma lógica *dedutiva*. Na medida em que as inferências do dia a dia não necessariamente preservam a verdade, uma lógica indutiva talvez pudesse capturá-las. Aliás, essa parece ser a conclusão dos psicólogos do assim chamado "novo paradigma", que menciono abaixo.

tro". Para fazer essa verificação, os participantes são solicitados a virar somente e todas as cartas que precisam ser viradas.

A ação correta, de acordo com a tabela de verdade da implicação material, seria virar apenas as cartas N e 8. Isso porque, se houver um número diferente de 6 do outro lado do N, a condicional será falsa; igualmente se houver um N do outro lado do 8. Ocorre que a maior parte dos participantes vira apenas a carta N, ou a carta N e a carta 6. "Este é um dos resultados mais robustos na história da psicologia do pensamento, se não da psicologia cognitiva em geral" (Elqayam, 2018, p. 134, minha tradução). E talvez isto ajude a explicar a sensação de artificialidade que estudantes costumam ter diante da implicação material.

A conclusão desses estudos é sintetizada por Elqayam (2018, p. 130, minha tradução): "quando a psicologia do pensamento emergiu como uma ciência experimental em meados do século XX, alguns dos seus resultados mais dramáticos mostraram que humanos não pensam logicamente; em outras palavras, a lógica revelou-se um sistema *descritivo* ruim". Quer dizer, sistemas lógicos dedutivos não são um bom modelo dos nossos raciocínios espontâneos.

Diante desses resultados experimentais, temos ao menos duas alternativas: ou assumimos que humanos, na maior parte do tempo, não são racionais; ou salvamos a racionalidade humana e aceitamos que sistemas lógicos dedutivos não capturam aquilo que faz dos humanos animais racionais. Os psicólogos têm cade vez mais migrado para a segunda opção.

No que Elqayam (2018) e Oaksford & Chater (2020) chamam de "novo paradigma" da psicologia do pensamento, os assim chamados "vieses cognitivos", tais como o viés de crença, não são mais considerados erros, mas características próprias de uma racionalidade majoritariamente probabilística. Neste novo paradigma, os raciocínios que efetuamos em situações cotidianas (e que pessoas não treinadas em lógica efetuam nos experimentos mencionados acima) são modelados pela teoria bayesiana de probabilidades. "O que aparenta ser uma resposta errada quando contrastado com cânones lógicos, em geral acaba se tornando racionalmente justificado quando visto pela noção mais rica de racionalidade do novo paradigma" (Oaksford & Chater, 2020, p.

305, minha tradução). Inferências probabilísticas precisam se basear em conhecimento prévio, dado que a avaliação de probabilidades demanda algum conhecimento do contexto sobre o qual se está fazendo inferências; daí o caráter não-monotônico das avaliações de silogismos feitas por pessoas não treinadas em lógica. Psicólogos adeptos do novo paradigma explicam de maneira similar os "erros" dos testados na tarefa de seleção de Wason (Oaksford & Chater, 1994).

Se os psicólogos do novo paradigma estão certos, então a hipótese racionalista não está. O objeto de estudo da lógica não pode ser a racionalidade humana (entendida em sentido amplo, como aquilo que caracteriza o "animal racional") simplesmente porque a racionalidade humana não é lógico-dedutiva.

7. Conclusão

Vimos como dados empíricos, advindos da psicologia, nos ajudam a descartar ou confirmar hipóteses filosóficas sobre o objeto de estudo da lógica, sem que para isso tenhamos que esposar qualquer forma de psicologismo. Aliás a conclusão a que esses próprios dados levam é francamente anti-psicologista: as leis lógicas não fazem parte da fisiologia da mente; elas precisam ser aprendidas, como os estudos com não-treinados em lógica sugerem. Vimos também que pelo menos um aspecto central das leis lógicas de primeira ordem têm origem pré-linguística: os conceitos de objeto e propriedade refletem aspectos estruturantes do modo como percebemos setores específicos da realidade macroscópica antes mesmo de adquirirmos linguagem. Mas é difícil ver essas categorias como estruturantes do próprio mundo como um todo. Há vastos setores do mundo que não parecem ser naturalmente estruturados em objetos. A aplicação universal dessas categorias, tal como sugerida pela lógica clássica, exige um elevado nível de idealização do mundo. Neste sentido, a estruturação universal do discurso e, por conseguinte, do domínio do discurso em objetos parece mais uma imposição ao mundo que o reflexo de um aspecto seu. A conclusão parece ser que sistemas lógicos dedutivos não refletem nos-

sos modos espontâneos de raciocinar, nem aspectos puramente linguísticos, nem aspectos da estrutura fundamental da realidade.

De onde vêm as leis lógicas, então? Uma resposta, que parece ser a de Frege, é que as leis lógicas são descrições verdadeiras de uma realidade não-espaçotemporal de relações lógicas. Esta resposta não agrada quem vê dificuldades com a admissão da existência de uma realidade não-espaçotemporal. Ademais, a resposta platonista não é suscetível a uma abordagem empírica, simplesmente porque tal realidade não-espaçotemporal está fora do alcance de qualquer método empírico.

Uma outra resposta para a origem da lógica, que é a que prefiro, é que as leis lógicas são regras estipuladas por nós com o objetivo de assegurar inferências que preservam necessariamente a verdade. Neste sentido, sistemas lógicos teriam origem em criações originais, inovações técnicas na arte de raciocinar. Eles capturariam, sim, alguns aspectos pré-existentes do nosso modo de pensar e perceber o mundo, como nossa tendência de organizar o mundo em objetos com propriedades e relações. Mas mesmo quando falamos e pensamos em termos de objetos com propriedades e relações bem definidas, não fazemos consistentemente inferências *dedutivas* sobre eles, como visto na seção 6. Ainda que racionais, nossas inferências probabilísticas nem sempre transmitem a verdade das premissas para a conclusão. Embora o ser humano não nasce naturalmente equipado com a habilidade de fazer inferências que necessariamente preservam a verdade, ele foi capaz de desenvolver um método para isso.

Pode ser que a reflexão crítica sobre processos de inferência espontâneos, como aqueles de pessoas não treinadas em lógica exemplificados na seção 6, tenha levado filósofos, lógicos e matemáticos do passado a concluir que nossas maneiras de pensar intuitivas são falhas quando o objetivo é transmitir necessariamente a verdade das premissas para a conclusão, incentivando-os a criar técnicas para assegurar esse grau de certeza. Estas técnicas tiram proveito da estruturação do domínio do discurso em objetos com propriedades e relações bem definidas, ainda que idealizadas. Filósofos e matemáticos criaram essas técnicas porque estavam em busca de certezas que seus modos espontâ-

neos de raciocínio não podiam garantir. O lugar primordial de emprego dessas técnicas tem sido, desde então, a ciência, a matemática e a filosofia. Como afirma Dutilh Novaes (2021, p. 20-21),

> o raciocínio dedutivo quase nunca é instanciado 'na natureza', por assim dizer, pois está em desacordo com o forte componente de derrotabilidade do raciocínio cotidiano. Na maioria das circunstâncias cotidianas, o raciocínio dedutivo passa do ponto: a questão não é inferir com absoluta certeza o que decorre *necessariamente* das informações disponíveis, mas sim o que é provável que aconteça, dadas as informações disponíveis e uma série de suposições. […] [O] raciocínio dedutivo pertence a nichos de especialistas: matemáticos, cientistas, filósofos, e mesmo nesses nichos ele não domina completamente outras formas de raciocínio.

Leis lógicas seriam estipulações primariamente normativas que não descrevem nenhuma realidade subjacente: elas ditariam como devemos raciocinar se o objetivo é fazer inferências que preservam a verdade, *contrariamente* ao que nossa natureza nos levaria a fazer, e *enquadrando* os contextos do mundo sobre os quais queremos fazer inferências dedutivas nas categorias idealizadas de objeto, propriedade e relação. Por certo, a hipótese de que leis lógicas sejam técnicas de inferência criadas por nós precisa ser defendida também com evidências históricas, que não apresentei aqui. Remeto o leitor interessado para Dutilh Novaes (2021), que mostra como maneiras de pensar dedutivas têm suas raízes em práticas dialógicas entre filósofos e matemáticos da Grécia Antiga.

Referências

ARENHART, Jonas. The Price of True Contradictions About the World. In: CARNIELLI, W.; MALINOWSKI, J. (eds). *Contradictions, from Consistency to Inconsistency*. Trends in Logic (Studia Logica Library), v. 47. Springer, 2018.

BALL, Linden. THOMPSON, Valerie. Belief Bias and Reasoning. In BALL, L.; THOMPSON, V. (eds.). *International handbook of thinking and reasoning*. New York, Routledge, 2018.

BÉZIAU, Jean-Yves. Logic is not logic. *Abstracta*, n. 6, v. 1, p. 73-102, 2010.

CARNAP, Rudolf. Empiricism, Semantics and Ontology. In BENACERRAF, P., & PUTNAM, H. (eds.). *Philosophy of Mathematics: selected readings* (p. 241-257). Cambridge, Cambridge University Press, 1983.

COHNITZ, Daniel. ESTRADA-GONZÁLES, Luis. An Introduction to the Philosophy of Logic. Cambridge, Cambridge University Press, 2019.

DA COSTA, Newton. *Ensaio sobre os fundamentos da Lógica*. São Paulo, Hucitec, 1994.

DA COSTA, Newton. ARENHART, Jonas. Full-blooded anti-exceptionalism about logic. *Australasian Journal of Logic*, n. 15, v. 2, p. 362-380, 2018.

DA COSTA, Newton. KRAUSE, Décio. Schrödinger Logics. *Studia Logica*, n. 53, p. 533-550, 1994.

DUTILH NOVAES, Catarina. *Formal Languages in Logic*: a philosophical and cognitive analysis. Cambridge, Cambridge Universtiy Press, 2012.

DUTILH NOVAES, Catarina. *The Dialogical Roots of Deduction*. Cambridge, Cambridge University Press, 2021.

ELQAYAM, Shira. The new paradigm in psychology of reasoning. In BALL, L.; THOMPSON, V. (eds.). *International handbook of thinking and reasoning*. New York, Routledge, 2018.

EVANS, J. BARSTON, Julie. POLLARD, Paul. On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, n. 11, v. 3, p. 295-306, 1983.

FODOR, J. A. *The language of thought*. New York: Thomas Y. Crowell Company, 1975.

FREGE, Gottlob. The Basic Laws of Arithmetic. Berkeley: University of California Press, 1964.

FREGE, Gottlob. Logic. In HERMES, H., KAMBARTEL, F., KAULBACH, F. (eds.). *Posthumous Writings*. Oxford: Basil Blackwell, 1979.

FRENCH, Steven. KRAUSE, Décio. Quantum vagueness. *Erkenntnis*, n. 59, p. 97-124, 2003.

HYLTON, Peter. *Quine*. New York, Routledge, 2007.

JOHNSON, Scott. Object Perception. In ZELAZO, P. (ed.). *The Oxford Handbook of Developmental Psychology*, v. 1. Oxford, Oxford University Press, 2013.

JOHNSON, Scott. Object Concept. In BENSON, J. (ed.). *Encyclopedia of Infant and Early Childhood Development*, 2nd edition. vol. 2. Elsevier, 2020.

KNUDSEN, Eric. Evolution of neural processing for visual perception in vertebrates. *The Journal of Comparative Neurology*, 2020.

KRAUSE, Décio. ARENHART, Jonas. Is Identity Really so Fundamental? *Foundations of Science*, 24, p. 51–71, 2019.

LUPYAN, Gary. THOMPSON-SCHILL, Sharon. SWINGLEY, Daniel. Conceptual Penetration of Visual Processing. *Psychological Science*, n. 21, v. 5, p. 682-691, 2010.

MACHERY, Édouard. *Philosophy Within Its Proper Bounds*. Oxford, Oxford University Press, 2017.

MADDY, Penelope. *Second Philosophy*: a naturalistic method. Oxford, Oxford University Press, 2007.

MENDES, Natacha. RAKOCZY, Hannes. CALL, Josep. Ape metaphysics: Object individuation without language. *Cognition*, n. 106, v. 2, 2008.

MERCIER, Hugo. Reasoning and Argumentation. In BALL, L.; THOMPSON, V. (eds.). *International handbook of thinking and reasoning*. New York, Routledge, 2018.

OAKSFORD, Mike. CHATER, Nick. A Rational Analysis of the Selection Task as Optimal Data Selection. *Psychological Review*, vol. 101, n. 4, p. 608-631, 1994.

OAKSFORD, Mike. CHATER, Nick. New Paradigms in the Psychology of Reasoning. *Annual Review of Psychology*, n. 71, p. 305-330, 2020.

PYLYSHYN, Zenon. Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, n. 22, p. 341-423, 1999.

QUINE, W.V.O. Two Dogmas of Empiricism. In Harding, S.G. (ed.). *Can Theories be Refuted?* Dordrecht, Springer, 1976.

RUSSELL, Gillian. Varieties of Logical Consequence by Their Resistance to Logical Nihilism. In WYATT, J., PEDERSEN, N., & KELLEN, N. (eds). *Pluralisms in Truth and Logic*. Cham (Switzerland), Palgrave Macmillan, 2018.

SÁ, Walter. WEST, Richard. STANOVICH, Keith. The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, 91(3), 497–510, 1999.

SCHLIESSER, Eric. Synthetic philosophy. *Biology & Philosophy*, *34*, v. 2, p. 1-9, 2019.

SORENSEN, Roy. Vagueness. In ZALTA, E. (ed.). *The Stanford Encyclopedia of Philosophy*, Summer edition, 2018.

TEUFEL, Christoph. NANAY, Bence. How to (and how not to) think about top-down influences on visual perception. *Consciousness and Cognition*, 2016.

VETTER, Petra. NEWEN, Albert. Varieties of cognitive penetration in visual perception. *Consciousness and Cognition*, n. 27, p. 62-75, 2014.

WILLIAMSON, Timothy. *Vagueness*. London, Routledge, 1994.

WILLIAMSON, Timothy. *Modal Logic as Metaphysics*. Oxford, Oxford University Press, 2013.

WILLIAMSON, Timothy. Semantic paradoxes and abductive methodology. In Armour-Garb, B. (ed.). *Reflections on the Liar*. Oxford, Oxford University Press, 2017.

CONTEÚDO EXISTENCIAL NAS LÓGICAS ARISTOTÉLICA E CLÁSSICA

Diogo Henrique Bispo Dias[1]

RESUMO

É comumente dito que a lógica aristotélica é incompatível com a lógica clássica, e o motivo para tal é o modo com o qual a lógica aristotélica lida com o conteúdo existencial de suas proposições. Segundo a posição padrão, Aristóteles pressupõe que todas as proposições categoriais têm conteúdo existencial e, sem tal pressuposto, a lógica aristotélica seria inconsistente. O objetivo deste artigo é mostrar, a partir da análise dos textos de Aristóteles, que esta tese é falsa. Argumentaremos que Aristóteles reconhecia a possibilidade de proposições categóricas com termos vazios, e que isso não afeta a consistência interna da lógica aristotélica. Por fim, mostraremos que, ainda assim, a lógica aristotélica continua sendo incompatível com a lógica clássica.

Palavras-chave: Conteúdo Existencial. Lógica Aristotélica. Quadrado da Oposição. Lógica Clássica.

ABSTRACT

It is commonly said that Aristotelian logic is incompatible with classical logic, and the reason for this is the way in which Aristotelian logic deals with the existential import of its propositions. According to the standard position, Aristotle assumes that all categorical propositions have existential import and, without such an assumption, Aristotelian logic would be inconsistent. The purpose of this article is to show, from an analysis of Aristotle's texts, that this thesis is false. We will argue that Aristotle recognized the possibility of categorical propositions with empty terms, and that this does not affect the internal consistency of Aristotelian logic. Finally, we will show that, even so, Aristotelian logic remains incompatible with classical logic.

Keywords: Existential Import. Aristotelian Logic. Square of Opposition. Classical Logic.

[1] Professor de Filosofia na Universidade Estadual do Norte do Paraná – UENP.
E-mail: diogo.bispo.dias@gmail.com. ORCID: 0000-0002-0601-6850.

1. Introdução

É comum se afirmar que a lógica aristotélica é incompatível com a lógica clássica. Fala-se isso, principalmente, por conta do modo com o qual a lógica aristotélica lida com o *conteúdo existencial*[2] de suas proposições.

Uma proposição tem conteúdo existencial quando ela pressupõe a existência de algum objeto. Assim, por exemplo, a sentença[3] "Algum homem é mortal" é geralmente considerada como possuindo conteúdo existencial, visto que ela parece pressupor a existência de, pelo menos, um homem.

A importância do conteúdo existencial para a lógica aristotélica é reconhecida por boa parte da história da filosofia. Ela foi discutida amplamente na Idade Média[4] e endossada por Leibniz[5]. No entanto, na passagem do século XIX para o século XX, com as primeiras tentativas de uma formalização matemática da lógica, foram evidenciados problemas com esse pressuposto.

Nesse momento, houve um enorme debate não somente sobre a compatibilidade da lógica aristotélica com a lógica matemática que estava surgindo, mas, também, sobre a própria consistência interna da lógica aristotélica. Muitos autores simplesmente defenderam que a lógica aristotélica era inconsistente, e que isso nunca tinha sido percebido ao longo da história da filosofia.

A posição padrão resultante deste debate é que todas as proposições da lógica aristotélica têm conteúdo existencial, e esta lógica só é consistente se aceitarmos este pressuposto. Por outro lado, como a lógica clássica não aceita este pressuposto, estas duas lógicas são, portanto, incompatíveis.

Nesse sentido, o objetivo desse texto é apresentar brevemente esse debate e mostrar que ainda que essas duas lógicas sejam, de fato, incompatíveis, o motivo desta incompatibilidade é mal interpretado por boa parte da literatura. Defenderemos, também, que há uma interpretação natural, e fiel a várias passagens dos textos de Aristóteles, que permite concluir que nem toda propo-

---

[2] Em inglês, usa-se a expressão "*existential import*". Também seria possível traduzi-la para o português como "importe existencial". Usamos "*conteúdo existencial*" para manter a terminologia presente nos principais manuais de lógica em português, como COPI (1978), e MORTARI (2017).

[3] Usaremos sentença e proposição como sinônimas.

[4] Cf. MOODY (1953).

[5] Cf. KATTSOFF (1934).

sição na lógica aristotélica pressupunha conteúdo existencial[6], e que esta interpretação torna esta lógica internamente consistente.

2. Lógica Aristotélica

Comecemos por alguns comentários gerais sobre a lógica aristotélica. Esta tem como noção central o conceito de dedução. Assim,

> Uma dedução é um discurso (*logos*) no qual, certas coisas sendo supostas, algo diferente das coisas supostas segue-se por necessidade. (ARISTÓTELES, Analíticos Anteriores I, 24b19–20)[7].

As coisas supostas são o que chamamos de *premissas*, e aquilo diferente que se segue destas é chamado de *conclusão.* A noção de necessidade presente nesta definição corresponde, em linhas gerais, à noção moderna de validade: um argumento é válido quando é impossível que suas premissas sejam verdadeiras, e a conclusão falsa. Ou ainda: a conclusão segue por necessidade das premissas se é impossível para a conclusão ser falsa quando as premissas são verdadeiras.

Nesta definição de dedução, já é possível identificar diferenças importantes com relação à noção moderna de validade. Dois tipos de argumentos permitidos na lógica clássica são, de princípio, excluídos da lógica aristotélica, a saber:

I. argumentos em que a premissa é igual à conclusão;
II. argumentos com uma única premissa.

Além disso, a lógica aristotélica lida somente com argumentos na forma chamada de silogismo categórico. Isto significa que as sentenças que fazem parte do silogismo são sentenças declarativas, ou seja, sentenças que podem ser consideradas verdadeiras ou falsas, nas seguintes formas:

---

[6] Além disso, o presente artigo se mostra importante, pois o debate sobre o papel do conteúdo existencial na lógica aristotélica é majoritariamente feito em língua inglesa.

[7] Todas as traduções para o português são do autor.

Sentença do tipo **A**: Todo S é P;

Sentença do tipo **E**: Nenhum S é P;

Sentença do tipo **I**: Algum S é P;

Sentença do tipo **O**: Algum S não é P.[8]

Aristóteles defende que toda asserção deve afirmar ou negar a atribuição de um único predicado a um único sujeito. Ou seja, a lógica aristotélica não permite, nos silogismos, sentenças conjuntivas ou disjuntivas, por exemplo. Podemos ver, também, que Aristóteles não considera negação como uma sentença composta.

Ainda assim, essas diferenças não são suficientes para afirmar que a lógica aristotélica é incompatível com a lógica clássica. Aquela poderia constituir um fragmento desta, e simplesmente diríamos que a lógica clássica aumentou o escopo de aplicação da lógica aristotélica. Para evidenciar a raiz dessa incompatibilidade, vejamos algumas relações entre proposições na lógica aristotélica:

I. Duas sentenças são *contraditórias* se é impossível que ambas sejam falsas, e ambas sejam verdadeiras.
II. Duas sentenças são *contrárias* se é impossível que ambas sejam verdadeiras, mas é possível que ambas sejam falsas[9].
III. Duas sentenças são *subcontrárias* se é impossível que ambas sejam falsas, mas é possível que sejam verdadeiras.
IV. Duas sentenças são chamadas de *subalterna* e *superalterna*, respectivamente, quando a verdade da superalterna implica a verdade da subalterna, e a falsidade da subalterna implica na falsidade da superalterna.

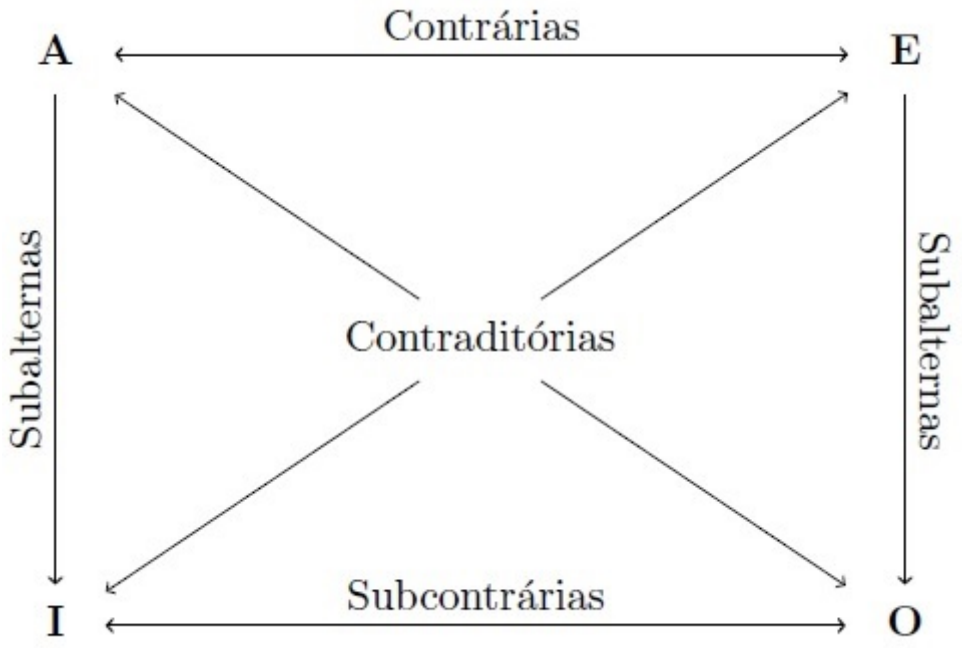
[8] Existe um debate acerca da presença – ou não – de sentenças singulares do tipo "Sócrates é mortal" na lógica aristotélica. As teses defendidas neste artigo são independentes do resultado desse debate. Sobre essa discussão, Cf. READ (2015).

[9] Para ver a formulação precisa dada por Aristóteles para as noções de contraditoriedade e contrariedade, Cf. ARISTÓTELES, Da Interpretação, 17b16-20, e 17b20-26, respectivamente.

Aristóteles afirma que **A** e **O** são contraditórias, que **E** e **I** são contraditórias[10], e que **A** e **E** são contrárias[11]. A partir destas afirmações, podemos inferir outras relações entre as sentenças. Para ver que **I** e **O** são subcontrárias, suponha que **I** é falsa. Logo, **E** é verdadeira. A contrária de **E**, **A**, é falsa. E sua contraditória, **O**, é verdadeira. Assim, **I** e **O** não podem ambas ser falsas. Por outro lado, é possível que tanto **I** quanto **O** sejam verdadeiras. Portanto, **I** e **O** são subcontrárias.

Por fim, se **A** é verdadeira, então **I** também o é. E, se **I** é falsa, isso significa que **A** também é falsa. Logo, uma sentença **A** é superalterna de sua sentença **I** correspondente[12], e essa sentença **I** é subalterna à sentença **A** correspondente. Um raciocínio análogo mostra que a relação de subalternação também se aplica a sentenças do Tipo **E** e **O**.

Essas relações podem ser apresentadas graficamente através do chamado **Quadrado da Oposição**:



Esse diagrama não foi formulado por Aristóteles e sua primeira aparição conhecida data do século II d.C, na obra de Apuleio de Madaura[13].

[10] Cf. ARISTÓTELES, Da Interpretação, 17b18-19.

[11] Cf. ARISTÓTELES, Da Interpretação. 17b6-8.

[12] Isto é, de uma sentença **I** com mesmo sujeito e predicado da sentença **A**.

[13] Cf. LONDEY e JOHANSON (1978), pp. 108-12.

Além das relações acima, Aristóteles afirma que **E** e **I** têm conversão simples[14], isto é, "Nenhum S é P" é equivalente em valor de verdade a "Nenhum P é S" e "Algum S é P" é equivalente em valor de verdade a "Algum P é S".

Assim, chamaremos de *lógica aristotélica* as relações constituintes do Quadrado da Oposição, a tese que de vale a conversão simples de **E** e **I** e a silogística aristotélica.

3. Críticas à lógica aristotélica

No início do século XX, com o desenvolvimento do que hoje é chamado de *lógica clássica*, vários autores começaram a apontar inconsistências na lógica aristotélica[15]. Uma das principais críticas, que se tornou padrão a partir da análise de Łukasiewicz[16] (1951), é a de que a lógica aristotélica não consegue lidar com a presença de termos vazios.

O argumento padrão apresentado é o seguinte: suponha que S seja um termo vazio. Logo, a sentença "Algum S é P" é falsa. Sua contraditória, "Nenhum S é P", portanto, é verdadeira. Assim, sua subalterna, "Algum S não é P" deve ser verdadeira. Mas, por hipótese inicial, S é um termo vazio. Algo está errado, pois a sentença "Algum S não é P" parece pressupor que existe, ao menos, um S.

Nessa situação, parece que praticamente todas as relações no quadrado das oposições se perdem. Se S é vazio, **I** e **O** podem ambas ser falsas, portanto, não são sentenças subcontrárias. Mas, neste caso, suas contraditórias, **A** e **E**, podem ser ambas verdadeiras. Perdemos, portanto, a relação de contrariedade também. Como **I** e **O** podem ser ambas falsas, e **A** e **E**, verdadeiras, perdem-se também as relações de subalternação.

Esses argumentos dependem de dois pressupostos:

---

[14] Cf. ARISTÓTELES, Analíticos Anteriores, 25a4-9.

[15] Cf. NORTHROP (1928), ŁUKASIEWICZ (1957), SMITH (2019), MIGNUCCI (2007), KEYNES (1887), ROSE (1968).

[16] Łukasiewicz é um dos precursores das lógicas não-clássicas, sendo o criador das primeiras lógicas multivaloradas (que admitem valores de verdade além de Verdadeiro e Falso), e responsável por escrever um artigo importante com a primeira crítica elaborada ao princípio de não-contradição. Cf. ŁUKASIEWICZ (2005).

I. É permitido o uso de termos vazios;
II. Proposições particulares têm conteúdo existencial, isto é, elas pressupõem a existência de um indivíduo que instancia seu sujeito. Assim, por exemplo, a sentença "Algum unicórnio é verde" é verdadeira quando existe ao menos um unicórnio, e este unicórnio é verde.

A partir desse tipo de raciocínio, Kneale e Kneale concluem que a lógica aristotélica "pode ser justificada, como um todo, se supusermos que ela estava lidando somente com termos que têm aplicação" (KNEALE & KNEALE, 1962, p. 59)[17], isto é, com termos que têm conteúdo existencial.

Mignucci argumenta que, além do termo sujeito, o termo predicado também deve ser não-vazio:

> De um ponto de vista técnico, para justificar a lógica aristotélica, em geral, precisamos evitar termos vazios não somente na posição do sujeito, mas, também, na posição do predicado. É verdade que, para evitar os contraexemplos ao quadrado da oposição é suficiente impor a exigência de que os sujeitos das proposições sejam termos não-vazios. Mas também é verdadeiro que Aristóteles também aceitava a conversão das proposições. […] Para aceitar essas teses, deve-se conceder que não somente sujeitos, mas, também, predicados devem ser termos não-vazios (MIGNUCCI, 2007, p. 137)[18].

Jacobs afirma que os complementos[19] de S e P também devem ser não-vazios:

> Para permitir todas as inferências imediatas de Aristóteles, devemos assumir que tanto os termos S e P, bem como seus complementos não-S e não-P são bem sucedidos em denotar (JACOBS, 1980, p. 419)[20].

---

[17] "[it] can be justified as a whole, if we suppose that [it] was dealing only with terms that have application".

[18] "From a technical point of view, to justify Aristotle's logic in general we need to avoid empty terms not only in a subjective but also in a predicative position. It is true that to avoid the counterexamples to the square of opposition it is sufficient to impose the requirement that the subjects of the propositions be not empty terms. But is also true that Aristotle accepted also conversion of propositions. […] To accept these views one must concede that not only subjects but also predicates must be non-empty terms".

[19] O complemento de uma classe é a coleção de todas as coisas que não pertencem a esta classe. Assim, o complemento da classe dos mortais é a classe de todas as coisas que são não-mortais.

[20] "In order to allow all of Aristotle's immediate inferences we must assume that both the terms S and P and their complements non-S and non-P successfully denote".

A partir destas considerações, tornou-se padrão aceitar que a lógica aristotélica só é consistente se todos os termos de todas as sentenças tiverem conteúdo existencial. Além disso, é comum dizer que essa pressuposição não era reconhecida por Aristóteles ou pelos medievais e que, portanto, todos estes faziam uso de uma lógica inconsistente.

Visto que a lógica aristotélica foi mantida por praticamente dois mil anos, é estranho pensar que tal falha passou incólume. Haveria outra explicação disponível? Mostraremos que sim. Como ficará claro, defenderemos que o quadrado da oposição é coerente com a presença de termos vazios. A partir da análise do próprio texto de Aristóteles, veremos que proposições negativas, **O** e **E**, não têm conteúdo existencial. Sendo assim, proposições desses tipos são verdadeiras por vacuidade se o sujeito é vazio.

Começaremos investigando o caso das sentenças afirmativas. A verdade da sentença "Algum homem é mortal" pressupõe, naturalmente, a existência de um homem. Portanto, é consenso que sentenças do tipo **I** têm conteúdo existencial.

Vejamos o caso de proposições afirmativas universais. Para tal, considere o seguinte silogismo[21]:

Todo C é B
Todo C é A
Algum A é B

Se considerarmos que sentenças do tipo **A** não têm conteúdo existencial, então o argumento é inválido. Mas como Aristóteles defende que o argumento é válido[22], sabemos que ele considera **A** como possuindo conteúdo existencial. Além disso, como a verdade de **A** implica na verdade de **I**, e **I** tem conteúdo existencial, **A** também deve ter.

Portanto, estabelecemos que as proposições categoriais afirmativas, particulares e universais, pressupõem conteúdo existencial. Até aqui, há um consenso na literatura. O ponto de discórdia está na interpretação da sentença

[21] Esta forma de silogismo é denominada, pelos medievais, de *Darapti*.
[22] Cf. ARISTÓTELES, Analíticos Anteriores I, 39a15-18.

particular negativa. A literatura padrão defende que Aristóteles aceitava que proposições do tipo **O** também têm conteúdo existencial. Mostraremos que esta tese é falsa[23].

Em primeiro lugar, é fácil ver que Aristóteles reconhece a existência de proposições com termos vazios, e a possibilidade de determinar o valor de verdade desse tipo de proposição[24]. Nas *Categorias*, Aristóteles afirma que:

> Com uma afirmação e negação, uma sempre será falsa e a outra verdadeira, quer [o sujeito] exista ou não. Pois tome "Sócrates é doente" e "Sócrates não é doente": se ele é doente, é claro que uma ou outra dessas sentenças será verdadeira ou falsa, e igualmente se ele não existe; pois, se ele não existe, "ele é doente" é falsa, mas "ele não é doente" é verdadeira. (ARISTÓTELES, Categorias, 13b23-27).

Aristóteles também frequentemente faz uso de termos vazios, como bode-cervo e centauros[25]. O próximo passo é verificar que algumas proposições categoriais também podem ter termos vazios. O segredo para entender que proposições negativas particulares não tem conteúdo existencial é reconhecer que, em Aristóteles, a forma **O** não é a usual "Algum S não é P", mas, sim, "Nem todo S é P". Como aponta Wedin[26], isto está no próprio texto de Aristóteles:

> Chamo uma afirmação e uma negação de opostos *contraditórios* quando aquilo que uma significa universalmente, a outra significa não-universalmente. Exemplo: 'todo homem é branco' – 'nem todo homem é branco' (οὐ πᾶς ἄνθρωπος λευκός)[27], 'nenhum homem é branco' – 'algum homem é branco' (ARISTÓTELES, Da Interpretação, 17b17-19)

[23] Esta tese está presente em THOMPSON (1953), MOODY (1953), PRIOR (1962), WEDIN (1978, 1990), PARSONS (2017) e READ (2015).

[24] Este argumento foi apresentado por PRIOR (1962), e desenvolvido por WEDIN (1978).

[25] Cf. ARISTÓTELES, Analíticos Anteriores I, 49a24 e Analíticos Posteriores I, 89b32, respectivamente.

[26] Cf. WEDIN (1990). Essa interpretação da sentença de tipo **O** já está em MOODY (1953), mas este estava simplesmente buscando formular uma interpretação coerente da lógica aristotélica, sem se preocupar se o próprio Aristóteles defenderia esta posição.

[27] Aqui, coloco a expressão original em grego para enfatizar a escolha por traduzir como "Nem todo homem é branco". O texto grego utilizado é ARISTÓTELES (1962). Agradeço à Amarilis Ferreira pela ajuda com o texto original.

Vemos o mesmo em: "o oposto de 'todo homem é branco' é 'nem todo homem é branco' (οὐ πᾶς ἄνθρωπος λευκός)" (ARISTÓTELES, Da interpretação, 18a5).

A formulação "Nem todo S é P" deriva da tradução do Aristóteles para o latim, feita por Boécio, que foi uma dos principais responsáveis pela popularização da obra de Aristóteles no ocidente. No entanto, nos seus próprios comentários à obra de Aristóteles, Boécio usa a forma "Algum S não é P". Parsons[28] conjectura que isso se deu pois Boécio considerava que as duas formas são equivalentes em latim. De qualquer modo, a leitura dos comentários de Boécio se popularizou, e virou a forma padrão de ler proposições do tipo **O**[29].

É relevante apontar que os comentários de Al-Farabi[30] sobre os *Analíticos Anteriores* apresenta frequentemente a proposição de tipo **O** como "Nem todo X é E". Curiosamente, o tradutor desta obra para o inglês escolhe traduzir este trecho como "Some X is not E [Algum X não é E]", e apenas reconhece em uma nota que não está sendo fiel ao texto original[31].

Os *Analíticos Anteriores* oferecem outra evidência da tese de que sentenças categoriais particulares negativas não têm conteúdo existencial. Nessa obra, Aristóteles também expressa sentenças do tipo **O** como "B não pertence a algum C" (τό τινι τῷ Γ τὸ Β μή ὑπάρχει), e afirma que esta sentença pode ser verdadeira em duas situações, "se [B] não pertence a nenhum [C], e se [B] não pertence a todo [C]" (ARISTÓTELES, Analíticos Anteriores I, 26b15-17). O ponto central, aqui, é que "é verdade que M não pertence a algum O mesmo se ele não pertence a nenhum O" (ARISTOTELES, Analíticos Anteriores I, 27b21-22).

A partir dessas considerações, podemos concluir que proposições do tipo **O** não têm conteúdo existencial. O último parágrafo também nos mostra que uma proposição do tipo **E** é verdadeira quando o sujeito é vazio, pois a segunda condição de verdade de "B não pertence a algum C", isto é, quando B

[28] Cf. PARSONS (2017).

[29] É curioso notar que, já no século XII, Abelardo criticava a formulação da sentença **O** presente nos comentário de Boécio. Cf. PARSONS (2017), nota 8.

[30] Al-Farabi é um filósofo Árabe do século IX. Cf. RESCHER (1963).

[31] Cf. RESCHER (1963), p. 61, nota 9.

pertence nenhum C, é a mesma condição de verdade da proposição universal negativa "Nenhum B é C".

Com essa leitura, considere que S é um termo vazio. Sendo assim, a sentença "Algum S é P" é falsa. Logo, "Nenhum S é P" é verdadeira e, portanto, "Nem todo S é P" é verdadeira e "Todo S é P" é falsa. Por outro lado, Se "Todo S é P" é verdadeira, então "Algum S é P" é verdadeira, pois proposições afirmativas têm conteúdo existencial. Logo, "Nem todo S é P" é falso, e, portanto, "Nenhum S é P" também é falso. Assim, temos todas as relações do quadrado preservadas.

4. Contraposição e obversão

Acima, consideramos como Lógica Aristotélica o Quadrado da Oposição, a silogística, mais a conversão simples. No entanto, há outros princípios que são geralmente considerados como parte desta lógica, a saber: contraposição e obversão. A contraposição consiste em substituir, em proposições categoriais do tipo **A** e **O**, o sujeito da proposição com o complemento do seu predicado, e substituir o predicado com o complemento do seu sujeito. Assim, a contrapositiva de "Todo homem é mortal" é "Todo não-mortal é não-homem". Obversão, por sua vez, trata-se da operação de mudar, em qualquer proposição categorial, a qualidade da proposição (de afirmativa para negativa, ou vice-versa), e trocar o predicado da proposição pelo complemento do predicado. Deste modo, a obsersão de "Todo homem é mortal" é Nenhum homem é não-mortal".

O problema é que, como aponta Parsons[32], a aceitação destes princípios está em conflito com a tese de que a lógica aristotélica é capaz de lidar com termos vazios. Para ver isto, suponha que "Todo homem é um ser" é verdadeira. Sua contrapositiva é "Todo não-ser é um não-homem". Como sentenças do tipo **A** têm conteúdo existencial, e o termo "não-ser" é vazio, a sentença "Todo não-ser é um não-homem" é falsa. Logo, partimos de uma sentença verdadeira e, através da contraposição, chegamos a uma sentença falsa.

[32] Cf. PARSONS (2017).

Este exemplo foi apresentado por Buridan para defender a ilegitimidade da contraposição. Ainda que, no século XII, vários autores aceitavam esta inferência[33], no século XIV, o problema apontado por Buridan era amplamente conhecido, e a contraposição era considerada inválida, a não ser quando os termos da proposição não eram vazios.

Com respeito à obversão, temos uma situação análoga. Suponha que "nenhum marciano é verde" é verdadeira, e que não existem marcianos. Logo, "todo marciano é não-verde" é falsa. Portanto, a obversão de **E** para **A** é ilegítima. Da mesma forma, se "nem todo marciano é verde" é verdadeira, e não existem marcianos, a sentença "algum marciano é não-verde" é falsa. Logo, a obversão de **O** para **I** é inválida.

É importante notar que o próprio Aristóteles não reconhece obversão como um procedimento geral. No entanto, ele reconhece algumas aplicações particulares:

> 'Nenhum homem é justo' segue-se de 'todo homem é não-justo', enquanto que sua oposta, 'nem todo homem é não-justo (οὐ πᾶς ἐστὶν ἄνθρωπος οὐ δίκαιος), segue-se de 'algum homem é justo' (pois deve existir um).[34] (ARISTÓTELES, Da Interpretação, 20a19-21)

Portanto, a obversão funciona para as sentenças de tipo **A** e **I**, isto é, sentenças afirmativas. Church afirma que Aristóteles nunca apresenta obversão de uma sentença negativa[35]. A partir das considerações acima, agora estamos em uma posição de entender esta ausência. Aristóteles não defende a obversão de proposições negativas, pois essa inferência não vale na presença de proposições sem conteúdo existencial.

Alguns medievais dos séculos XII e XIII defendiam a legitimidade da obversão, e outros a rejeitavam. No século XIV, novamente, era claro que a obversão de sentenças negativas só valia quando os termos da proposição não eram vazios[36].

---

[33] Cf. PARSONS (2017).

[34] Note que, aqui, o conteúdo existencial da sentença de tipo **I** é explícito.

[35] Cf. CHURCH (1965), p. 419.

[36] Cf. PARSONS (2017), notas 16 e 17.

5. Objeções

Alguns autores rejeitam essa solução proposta. Priest[37] (2003), por exemplo, diz que aceitar que a sentença "Algum S não é P" não tem conteúdo existencial é estranho[38]. Ademais, ele reconhece que essa aceitação invalidaria a obversão, pois podemos passar de "Nenhum A é B" para "Todo A é não-B". Assim, poderíamos inferir conteúdo existencial de uma proposição sem conteúdo existencial.

A primeira objeção é adequadamente dissolvida quando reconhecemos que a melhor forma de formular uma sentença do tipo **O** é "Nem todo S é P". Ainda que seja estranho aceitar que a proposição "Algum marciano não é verde" possa ser verdadeira se não existe nenhum marciano, o mesmo não acontece se considerarmos a proposição "Nem todo marciano é verde". Esta sentença só será falsa se todos os marcianos forem verdes. Assim, se não existirem marcianos, "nem todo marciano é verde" é verdadeira por vacuidade.

Por sua vez, a segunda objeção é adequadamente rejeitada a partir dos comentários sobre obversão da seção anterior. Priest está correto em afirmar que essa interpretação invalidaria a obversão de **E** para **A**. O erro é não reconhecer que tanto Aristóteles, quanto os principais lógicos medievais sabiam disso e, exatamente por isso, ou a obversão não era defendida, ou era explicitado que ela dependia da postulação de conteúdo existencial das sentenças envolvidas.

Em suma, sentenças afirmativas contêm conteúdo existencial, ao passo que sentenças negativas não contêm. O argumento contra o quadrado da oposição não foi levantado antes do século XIX, pois essa era uma visão padrão. Por exemplo, Ockham afirma que:

> Em proposições afirmativas, sempre se declara que um termo supõe algo. Logo, se ele não supõe nada, a proposição é falsa. No entanto, em proposições negativas, a declaração é de que ou o termo não supõe nada, ou ele supõe algo sobre o qual o predicado é verdadeiramente negado. Assim, uma proposição negativa tem duas causas de verdade (OCKHAM, 1974, p. 206)[39].

[37] Cf. PRIEST (2003).

[38] O mesmo ponto é defendido por MIGNUCCI (2007), pp. 127-8.

[39] "In affirmative propositions a term is always asserted to supposit for something. Thus, if it supposits for nothing the proposition is false. However, in negative propositions the asserti-

Essa mesma posição é reafirmada, no século XIV, por Buridan, Ockham e Paulo de Venice.

No século XIX, os principais livros de lógica de língua inglesa incluíam contraposição e obversão como inferências legítimas, mas sem nenhuma discussão profunda sobre o papel do conteúdo existencial[40]. As obras de Łukasiewicz (1951) e de Kneale e Kneale (1962) atribuíram a Aristóteles essa visão da lógica desenvolvida ao longo do século XIX, que se torna inconsistente na presença de termos vazios.

No entanto, como vimos, esta crítica não se aplica ao que chamamos de *lógica aristótélica*, isto é, à lógica explicitamente endossada por Aristóteles. Alguns autores alertavam para essa confusão, reconhecendo que esta acusação de inconsistência à lógica aristotélica "dependia do pressuposto, não feito na lógica medieval, de que o conteúdo existencial deve ser atribuído às formas negativas assim como às afirmativas" (MOODY, 1953, p. 53)[41]. Do mesmo modo, Ashworth lamenta que

> A maioria dos lógicos depois da terceira década do século XVI parou de discutir esses assuntos [i.e., conteúdo existencial das proposições], com o resultado de que os leitores modernos tendem a pensar que a lógica tradicional carece de uma sofisticação que ela, de fato, possui (ASHWORTH, 1973, p. 147)[42].

6. Versão Moderna do Quadrado:

Chegamos à conclusão de que a interpretação adequada da lógica aristotélica é afirmar que proposições **A** e **I** têm conteúdo existencial, mas proposições **E** e **O**, não. Veremos, agora, que, mesmo com essa interpretação, a lógica aristotélica continua incompatível com a lógica clássica.

As proposições categóricas da lógica aristotélica são formalizadas, na lógica clássica, como:

---

on is either that the term does not supposit for something or that it supposits for something of which the predicate is truly denied. Thus a negative proposition has two causes of truth".

[40] Cf. PARSONS (2017), nota 25.

[41] "rest[s] on the assumption, not made in mediaeval logic, that existential import is to be assigned to the negative forms as well as the affirmative ones".

[42] "The vast majority of logicians after the third decade of the 16th century ceased to discuss these matters, with the result that modern readers tend to think of traditional logic as lacking a sophistication which it did indeed possess".

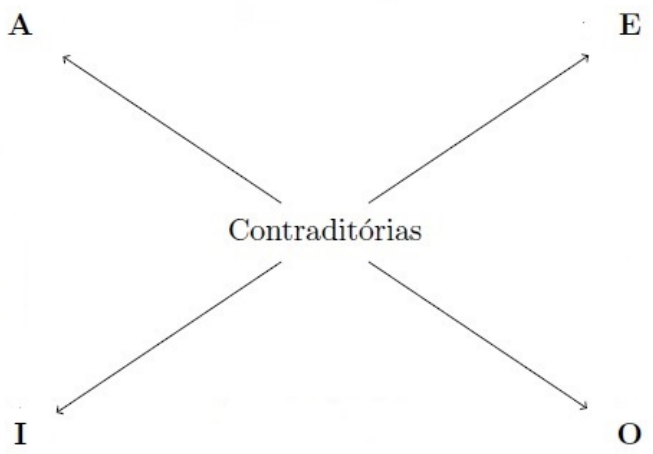
Todo S é P	$\forall x\ (Sx \rightarrow Px)$
Nenhum S é P	$\forall x\ (Sx \rightarrow \neg Px)$
Algum S é P	$\exists x\ (Sx\ \&\ Px)$
Algum S não é P	$\exists x\ (Sx\ \&\ \neg Px)$

Para ver que as duas lógicas são incompatíveis, basta notar que o silogismo *Darapti*:

Todo C é B
Todo C é A
Algum A é B

é válido na lógica aristotélica, mas inválido na lógica clássica. O motivo para isso é óbvio. Todas as sentenças afirmativas têm conteúdo existencial na lógica aristotélica. A lógica clássica, por sua vez, relaciona o conteúdo existencial com a quantidade das proposições categoriais, e não com a qualidade. Assim, proposições universais não têm conteúdo existencial, e proposições particulares têm. Sendo assim, o silogismo acima se torna inválido quando permitimos que C seja vazio.

Com essa reformulação, as relações do quadrado das oposições praticamente somem:



Suponha que S é vazio. Assim, **A** e **E** são ambas verdadeiras e, portanto, não são contrárias. Por sua vez, **I** e **O**, por terem conteúdo existencial na lógica clássica, podem ser ambas falsas e, logo, não são subcontrárias. Como **A** e **E** podem ser verdadeiras, e **I** e **O**, falsas, também perdemos as relações de subalternação. Portanto, considerando que apenas as sentenças particulares têm conteúdo existencial, as únicas relações que sobram no quadrado moderno são as de contraditoriedade. Segundo Church, essa visão da lógica aristotélica surgiu no final do século XIX, nas obras de Cayley – ainda que implicitamente –, Brentano, Peirce e Venn[43].

É curioso que esta interpretação seja frequentemente chamada de "interpretação booleana", dando a impressão de que teria sido formulada por George Boole. De fato, o trabalho de Boole forneceu o aparato formal no qual o papel do conteúdo existencial das proposições poderia ser melhor explicitado. No entanto, Boole nunca levou a cabo tal explicitação, tampouco estava interessado em comparar sua formulação com a obra de Aristóteles. Em particular, Boole aceitava a inferência de uma proposição do tipo **A** para uma proposição do tipo **I**[44] e, portanto, aceitava que proposições universais tinham conteúdo existencial. A discussão sobre o conteúdo existencial das proposições categóricas e a tese de que apenas sentenças particulares continham tal conteúdo só se tornaram explícitas com Brentano e, principalmente, com a formulação dos diagramas de Venn. Por esse motivo, Prior chama esta interpretação de "interpretação de Brentano-Venn"[45].

Logo, a lógica aristotélica não é compatível com a lógica clássica. Sua compatibilidade só se torna possível ao mudarmos o pressuposto aristotélico fundamental de que o conteúdo existencial é determinado pela qualidade das proposições categóricas, e não por sua quantidade.

---

[43] Cf. CHURCH (1965), p. 422.

[44] Cf. BOOLE (1847), p. 38. Para uma análise da lógica algébrica de Boole, evidenciando que esta lógica ainda pressupunha que proposições universais contêm conteúdo existencial, cf. VALENCIA (2004).

[45] Cf. PRIOR (1976), p. 123. Para uma análise do desenvolvimento do debate sobre conteúdo existencial no século XIX, cf. WU (1969).

7. Conclusão

Esse artigo apresentou uma interpretação minoritária da lógica aristotélica. Uma das consequências do não reconhecimento desse debate acerca do conteúdo existencial é que os principais livros didáticos de lógica simplesmente se limitam a repetir a tese que se consolidou, a partir do século XIX, de que a lógica aristotélica não é capaz de lidar com termos vazios. Por exemplo, Hausman, Kahane e Tidman afirmam que "a lógica silogística tradicionalmente depende do pressuposto de que todas as proposições [que figuram nos silogismos] têm conteúdo existencial" (HAUSMAN, KAHANE & TIDMAN, 2007, p. 319)[46]. Mortari repete a tese de que

> O conflito [entre as lógicas aristotélica e clássica] tem a ver com o fato de que, do ponto de vista contemporâneo, proposições universais não têm conteúdo existencial, ao contrário das particulares. Para a interpretação tradicional dos silogismos, contudo, proposições universais também tinham conteúdo existencial[47]. (MORTARI, 2017, p. 521)

Copi, ademais, enfatiza que é preciso estender essa exigência aos complementos dos termos:

> Se fizermos a pressuposição geral de que todas as classes designadas pelos nossos termos (e seus complementos) têm membros, então [...] a totalidade das relações estabelecidas no Quadro de Oposição tradicional mantém-se de pé (COPI, 1978, p. 157).

Como vimos, nenhuma destas teses está em Aristóteles, e as implicações da aceitação ou rejeição de termos vazios foi amplamente reconhecida e discutida pelos lógicos medievais.

Assim, esse artigo pretende contribuir, por um lado, para uma melhor compreensão dos pressupostos filosóficos envolvidos na lógica aristotélica, e o modo como esses pressupostos foram interpretados ao longo da história. Por outro lado, esperamos que este tipo de investigação mostre o papel fundamental dos aspectos filosóficos envolvidos na formulação de uma lógica, e de como esses aspectos influenciam em questões que geralmente são consideradas

[46] "Syllogistic logic traditionally rested on the assumption that all the propositions to be dealt with do have existential import".

[47] O autor também pressupõe que proposições particulares na lógica aristotélica têm conteúdo existencial.

meramente técnicas, como, por exemplo, determinar a validade de um argumento[48].

Referências

ARISTOTLE. *The Categories. On Interpretation*. Cambridge: Harvard University Press, 1962.

ASHWORTH, E. J. "Existential assumptions in late medieval logic". *American Philosophical Quarterly,* 10, 1973, pp. 141–147.

BARNES, J. *The Complete Work of Aristotle.* Vol. 1. Princeton: Princeton University Press, 1991.

BOOLE, G. *The Mathematical Analysis of Logic*. Cambridge: Cambridge University Press, 1847.

CHURCH, A. "The History of the Question of Existential Import of Categorical Propositions". In: BAR-HILLEL, B. (ed.). *Logic, Methodology, and Philosophy of Science* (Proceedings of the 1964 International Congress). Amsterdam: North-Holland, pp. 417–24.

COPI, I. *Introdução à Lógica*. 2ª Edição. São Paulo: Mestre Jou, 1978.

HAUSMAN, A.; KAHANE, H.; TIDMAN, P. *Logic and Philosophy*. Tenth Edition. Belmont: Thomson Wadsworth, 2007.

KATTSOFF, L. "Concerning the Validity of Aristotelian Logic". *Philosophy of Science*. Vol. 1, nº. 2, 1934, pp. 149-162.

KEYNES, J. N. *Studies and Exercises in Formal Logic*. Cambridge: McMillan and Co., 1887.

ŁUKASIEWICZ, J. *Aristotle's Syllogistic from the Standpoint of Modern Logic*. Oxford: Clarendon Press, 1951.

_____. "Sobre a Lei da Contradição em Aristóteles". In: ZINGANO, M. (org.). *Sobre a Metafísica de Aristóteles.* São Paulo: Odysseus, 2005, pp. 1-24.

MATES, B. *Elementary Logic*. New York: Oxford University Press, 1972.

MIGNUCCI, M. "Aristotle on the Existential Import of Propositions". *Phronesis*, Vol. 52, nº 2, 2007, pp, 121-138.

[48] Gostaríamos de agradecer o(a) parecerista anônimo(a) pela leitura cuidadosa do texto, bem como pelas valiosas sugestões.

MOODY, E. A. *Truth and Consequence in Mediaeval Logic*. Amsterdam: North-Holland, 1953.

MORTARI, C. A.; *Introdução à Lógica*. 2ª edição. São Paulo: Unesp, 2017.

NORTHROP, F. S. C. "An Internal Inconsistency in Aristotelian Logic". *The Monist*. Vol. 38, nº 2, 1928, pp. 193-210.

OCKHAM, W. *Summa Logicae: Theory of Terms. Part I*. Indiana: University of Notre Dame Press, 1974.

PARSONS, T. "The Traditional Square of Opposition". In: ZALTA, E.N. (ed.). *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition). Disponível em https://plato.stanford.edu/archives/sum2017/entries/square/. Acesso em: 20 set. 2020.

PRIEST, G. "On Alternative Geometries, Arithmetics, and Logics: a Tribute to Łukasiewicz". *Studia Logica*, 74, 2003, pp. 441-468.

PRIOR, A. N. *Formal Logic*. Oxford: Oxford University Press, 1962.

_____. *The Doctrine of Propositions and Terms*. London: Duckworth, 1976.

READ, S. "Aristotle and Łukasiewicz on Existential Import". *Journal of the American Philosophical Association*, Vol. 1, Nº3, 2015, pp. 535-44.

RESCHER, N. *Al-Farabi's Short Commentary on Aristotle's Prior Analytics*. Pittsburgh: University of Pittsburgh Press, 1963.

ROSE, L. E. *Aristotle's Syllogistic*. Illinois: Charles C. Thomas, 1968.

SMITH, R. "Aristotle's Logic". In: ZALTA, E. N. (ed.). *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Disponível em:

https://plato.stanford.edu/archives/sum2019/entries/aristotle-logic/. Acesso em: 20 set. 2020.

VALENCIA, V. S. "The Algebra of Logic". In: GABBAY, D. M.; WOODS, J. H. (eds.). *Handbook of the History of Logic. Rise of Modern Logic: From Leibniz to Frege.* Amsterdam: North-Holland, 2004, pp. 389-544.

WEDIN, M. V. "Aristotle on the existential import of singular sentences". *Phronesis,* 23, 1978, pp. 179-196.

_____. "Negation and Quantification in Aristotle". *History and Philosophy of Logic*, 11:2, 1990, pp. 131-150.

WU, J. S. “The Problem of Existential Import (From George Boole to P. F. Strawson)”. *Notre Dame Journal of Formal Logic*, vol. X, n. 4, 1969, pp. 415-424.

OPPOSITIONS IN A POINT

Alexandre Costa-Leite[1]

ABSTRACT

Following a previous article (cf. (Costa-Leite 2018)) in which logical oppositions are defined in a line segment, this article goes one step further and proposes a method to organize them using a zero-dimensional object: a point.

Keywords: Logical Oppositions. Diagrams. Dimension. Point.

RESUMO

Seguindo um artigo anterior (cf. (Costa-Leite 2018)) no qual as oposições lógicas são definidas em um segmento de reta, este artigo vai um passo além e propõe um método para organizá-las usando um objeto de dimensão zero: um ponto.

Palavras-chave: Oposições Lógicas. Diagramas. Dimensão. Ponto.

Introduction

In the environment of a two-valued logic such as first-order logic, (logical) oppositions are regularly four: contradiction, contrariety, subcontrariety and subalternation. By the very beginning, these relations were displayed using the square of opposition and, later, Robert Blanché in (Blanché 1966) showed how the square induces another bidimensional object: the hexagon of opposition. Recently, some authors investigated generalizations of the square to hexagons and to three-dimensional solids and so on (cf. (Beziau 2012), (Moretti 2004) and (Moretti 2012)).

In (Costa-Leite 2018), it is argued that a line segment of integers is enough to organize logical oppositions in such a way that there is no need to use n-dimensional objects ($n \geq 2$) to get representations of standard logical oppositions.[2] The procedure proposed works for the square of opposition and, with some adaptations, works as well for the hexagon of opposition. New con-

[1] Department of Philosophy, University of Brasilia, Brazil.
E-mail: costaleite@unb.br. ORCID: 0000-0002-5392-2186.

nections with respect to the line segment have been established by Fabien Schang in (Schang 2018).

Currently, a different direction is pursued but taking into account the results of (Costa-Leite 2018). Instead of a one-dimensional object such as a line segment of integers, this paper represents oppositions in an abstract dot, a point, a zero-dimensional object. Therefore, there is no need to use any dimension at all to arrange logical oppositions.[3]

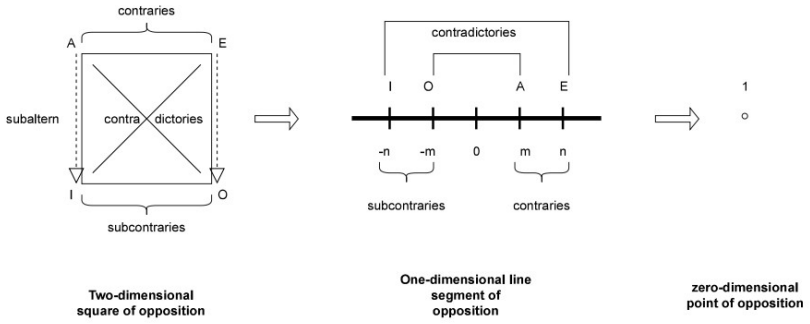


Figure 1: Reducing dimensions of diagrams

Figure 1 reflects the scheme of reductions concerning dimensions of diagrams used to represent oppositions. The square is initially transformed into a line segment which, in its turn, is transformed into a mathematical point.

Oppositions without dimensions

In (Costa-Leite 2018), the set $\mathbb{Z}$ is taken into consideration and categorical statements are associated to positive and negative integers allowing op-

[2] It is useful to know that some ontological economy can be produced also in the study of the relationships between n-dimensional ($n \geq 0$) diagrams and logical oppositions, as a debt with Ockham's razor, at least.

[3] When I developed the *basic construction* to convert the square and the hexagon into a line segment, a one-dimensional structure, Jean-Yves Béziau asked to me (in personal communication, 2019) whether it would be possible to arrange oppositions without dimensions. This article is a reply to his question. Thanks to Edelcio de Souza, Rodrigo Freire and Peter Arndt for comments and discussions. Thanks also to the reviewer (anonymous) for providing polished comments on the arguments in this paper.

positions to be established in a line segment. This produces, therefore, a conversion of the square and the hexagon of opposition into a line segment of opposition (for precision and details cf. (Costa-Leite 2018)).

In order to get started, consider classical two-valued first-order logic.[4] The following strategy defines all four oppositions by means of a mathematical dot, a point, a fragment of zero-dimensional space: it does not work, however, for all points in general. Differently, it works precisely for an unique and special point in the real line and the reasons for this fact will be clear in due time.

The square in a point

Let $exp(a)$ denotes the exponent of a (the construction uses physical distinctions between exponents with the same base). Again, as in (Costa-Leite 2018), let $\mathbf{C} = \{A, E, I, O\}$ be a set of categorical-like sentences containing statements of the form *A*, *E*, *I* and *O* (these are forms of sentences in a square of opposition), and for $m \neq n$ and $m, n \neq 0$, let $\mathbb{Z}' = \{-n, -m, m, n\} \subset \mathbb{Z}$ be a set of integers used to generate exponents of a given number a. Let $a \in \mathbb{Z}$ and let $\mathbb{Z}^\sharp$ be a set of the following type: $\mathbb{Z}^\sharp = \{a^{-n}, a^{-m}, a^m, a^n\}$. Now, consider a bijection $e : C \rightarrow \mathbb{Z}^\sharp$ such that $e(A) = a^m$, $e(E) = a^n$, $e(I) = a^{-n}$ and $e(O) = a^{-m}$. Thus, each member of $\mathbf{C}$ is assigned to a member of $\mathbb{Z}^\sharp$ in the following way: positive exponents are connected with universal statements and negative exponents are connected with existential statements. Let α, β be variables for elements of $\mathbf{C}$. So, $e(\alpha)$ and $e(\beta)$ are numbers of the form a^j, for $j \in \mathbb{Z}'$ while $exp(e(\alpha))$ and $exp(e(\beta))$ refer to the exponents of a.

Logical oppositions have to be redefined in this new setting according to the laws below (assuming that oppositions are mutually exclusive):

1. α and β are *contradictories* if, and only if $exp(e(\alpha)) + exp(e(\beta)) = 0$;
2. α and β are *contraries* if, and only if $exp(e(\alpha))$ and $exp(e(\beta))$ are positive integers;

[4] In an *n*-valued logic, with $n > 2$, other oppositions could appear.

3. α and β are *subcontraries* if, and only if $exp(e(\alpha))$ and $exp(e(\beta))$ are negative integers;
4. β is subaltern of α if, and only if, $exp(e(\beta)) \neq -exp(e(\alpha))$ and $exp(e(\beta))$ is a negative integer.

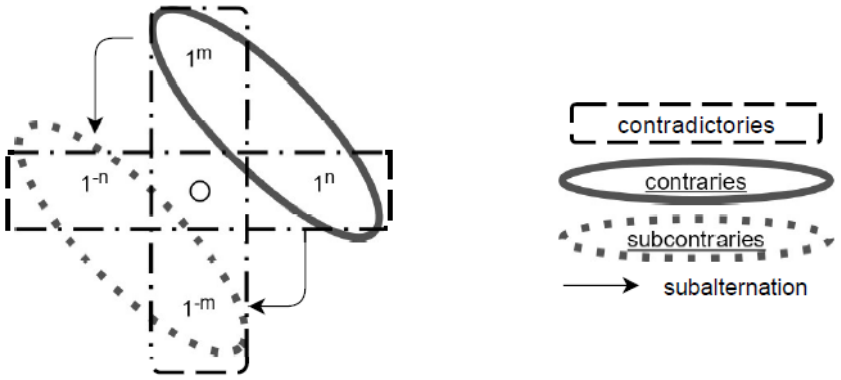


Figure 2: Oppositions in a zero-dimensional object

It is easy to see that the above construction does not lead to a zero-dimensional object for all a such that $a \in \mathbb{Z}^*$.[5] Take, for instance, $a = 2$ and let $\mathbb{Z}' = \{-2, -1, 1, 2\}$. In this case, the function e generates: $e(A) = 2^1$, $e(E) = 2^2$, $e(I) = 2^{-2}$ and $e(O) = 2^{-1}$. Despite the fact that all oppositions are satisfied, considering $a = 2$, this induces a line segment in the interval $[1/4, 4]$. Indeed, more generally, every time there is an $a \in \mathbb{Z}$ such that $a > 1$ and the greatest $exp(a)$ is k, then a line segment of the form $[1/a^k, a^k]$ is obtained.

So, in order to get oppositions in a point, we need a special point a in $\mathbb{Z}$. This is the number 1, the *surprising point*:[6] the unique point, in this method, in which all oppositions can be defined without leading to a line segment.[7] Let's check why. Consider $a = 1$ and let $\mathbb{Z}' = \{-2, -1, 1, 2\}$. Therefore, the function e associates $e(A) = 1^1$, $e(E) = 1^2$, $e(I) = 1^{-2}$ and $e(O) = 1^{-1}$. So, in

[5] We always take $a \neq 0$ because exponentiation is defined with base a.

[6] The problem of whether there is another construction allowing all points, in the real line or some fragment of it, to simulate oppositions in a zero-dimensional perspective should be examined.

[7] Notice that if we take into consideration 0 and its multiplications by numbers in a set of the form $\mathbb{Z}' = \{-n, -m, m, n\}$, using oppositions as above with obvious adaptations for multiplication, then 0 is also a point in which oppositions can be obtained without dimension. Thanks to Rodrigo Freire for this remark.

all cases, we have the number 1 and oppositions precisely setted in it. *A* and *O* (and *E* and *I*) are *contradictories* because the $exp(e(A)) + exp(e(O)) = 0$ ($exp(e(E)) + exp(e(I)) = 0$). We also have that *A* and *E* are *contraries* because that $exp(e(A))$ and $exp(e(E))$ are positive integers. In the same way, *I* and *O* are *subcontraries* as $exp(e(I))$ and $exp(e(O))$ are negative integers. Finally, notice that *I* is subaltern of *A* given that $exp(e(I)) \neq -exp(e(A))$ and $exp(e(I))$ is a negative integer. The same holds for *E* and *O*. Figure 2 represents how oppositions can be decorated in the point 1 (in the center) with its different manifestations using different exponents.[8]

Mutatis mutandis, following the ideas proposed in (Costa-Leite 2018) combined with the technique above, the two-dimensional hexagon of opposition can also be converted into a point, and this is let as an exercise to the reader. So, both the square and the hexagon can be transformed step-by-step into a line segment and into a point.

Conclusion

The theory of logical oppositions, as it has been developed in the literature, uses *n*-dimensional diagrams ($n \geq 2$) to represent contradiction, contrariety, subcontrariety and subalternation. Nevertheless, it is possible to represent oppositions using a line segment (one-dimensional space) and, as showed here, even no dimension at all is required, considering that oppositions can be defined in the surprising point 1.

A criticism to this approach has been raised by Peter Arndt from the University of Düsseldorf (in personal communication, 2020). He argues that, despite the fact that only the number 1 is used to define oppositions, the one-dimensional character of the construction appears in the exponents. This restores, therefore, one-dimensionality. It is somewhat reasonable to accept that there is, indeed, a parasitary connection between oppositions without dimensions and those organized in a line segment. The worth and essential remark here is that all representations used of the number 1 refer to the same object (1^m

[8] Zooming in the third diagram of figure 1 (in which the zero-dimensional point of opposition is represented) we see what is displayed in figure 2. The point 1 in the center shows that the diagram is zero-dimensional, and, let's say, it has one-dimensional peripheral orbits.

$= 1^n = 1^{-m} = 1^{-n} = 1$) and logical oppositions make use of only a single mathematical dot to formulate all oppositions in a zero-dimensional diagram. The construction developed shows the zero-dimensional point of opposition and it is dependent of a line segment in the exponents. However, what is relevant is the dimension of the diagram, and in the case of the point of opposition, it is its base which tells us in which dimension the diagram is, not the exponents. So, it is legitimate to speak about oppositions without dimensions. A next step would be to find a zero-dimensional point of opposition without any kind of dependence. Moreover, it is important to say that the line segment suggested in (Costa-Leite 2018) and the method here are not unique ways to reduce dimensions of diagrams, maybe some other possible reductions could also be achieved.

Two-dimensional or three-dimensional diagrams to represent oppositions are very well-known, and they have been used widely in the literature (especially squares). The work of Alessio Moretti in (Moretti 2012) contains the foundations of what is called *n*-opposition theory with some exploitations concerning dimensions of simplexes, but there is no attempt there to define all oppositions simultaneously in a given simplex of dimension one or zero. Again, concerning the plurality of dimensions, a natural ontological point is to check whether Ockham's razor should be applied in the case of diagrams. Why should we proliferate dimensions using complicated diagrams if we can use a line or a point? Is there any special reason for this fact? The existence of simple diagrammatic abbreviations such as a *line segment of opposition* and a *point of opposition* challenges the limits of the researches on the square and its extensions.

References

Beziau, J-Y. (2012). The power of the hexagon. *Logica Universalis*, 6(1), pp.1-43.

Blanché, R. (1966). *Structures intellectuelles. Essai sur l'organisation systématique des concepts*. Paris: Vrin.

Costa-Leite, A. (2018). Oppositions in a line segment. *South American Journal of Logic*, 4(1), pp.185-193.

Moretti, A. (2004). Geometry of Modalities? Yes: Through *n*-Opposition Theory. In: *Aspects of Universal Logic*, J-Y.Beziau, A. Costa-Leite, A.Facchini (editors), *Travaux de Logique*, 17, pp. 102-145.

Moretti, A. (2012). From the "Logical Square" to the "Logical Poly-Simplexes." In: *The Square of Opposition: a general framework for cognition*. J-Y. Beziau and G. Payette (editors), pp. 119-156. Bern: Peter Lang.

Schang, F. (2018). End of the Square? *South American Journal of Logic*, 4(2), pp.485-505.

ARITMÉTICA E CONHECIMENTO SIMBÓLICO: NOTAS SOBRE O *TRACTATUS LOGICO-PHILOSOPHICUS* E O ENSINO DE FILOSOFIA DA MATEMÁTICA

Gisele Dalva Secco[1]

À memória de Sören Stenlund

RESUMO

Partindo de e encerrando com reflexões acerca de práticas de ensino de filosofia da matemática, proponho uma comparação entre os principais traços da noção leibniziana de conhecimento simbólico e algumas passagens tractarianas sobre a aritmética. Defendo que esta chave de leitura permite a um só tempo (i) projetar nova luz sobre as especificidades da definição tractariana de número em comparação com as de Frege e Russell; (ii) fazer despontar a compreensão da natureza do conhecimento matemático como conhecimento simbólico ou formal que Wittgenstein mobiliza em seu livro; (iii) elencar algumas razões para a alegação de que Wittgenstein pode ser considerado o filósofo da prática matemática *avant la lettre*. O trabalho se encerra com um apanhado, um retorno à reflexão inicial sobre os vínculos entre pesquisa e ensino, e uma defesa da chave de leitura aqui utilizada em termos de seu potencial para o desenvolvimento de estudos em filosofia da matemática.

Palavras-chave: Conhecimento simbólico. Leibniz. Wittgenstein. Ensino de Filosofia. Filosofia da prática matemática.

ABSTRACT

Departing from and closing with reflections on issues regarding teaching practices of philosophy of mathematics, I propose a comparison between the main features of the Leibnizian notion of symbolic knowledge and some passages from the *Tractatus* on arithmetic. I argue that this way of reading allows (i) to shed a new light on the specificities of the Tractarian definition of number, compared to those of Frege and Russell; (ii) to highlight the understanding of the nature of mathematical knowledge as symbolic or formal knowledge that Wittgenstein mobilizes in his book; (iii) to offer reasons for the claim that Wittgenstein can be considered the philosopher of mathematical practice *avant la lettre*. The paper ends with an overview, a return to the initial reflection on the connections between research and teaching, and a de-

[1] Professora adjunta do Departamento de Filosofia da Universidade Federal de Santa Maria (UFSM). E-Mail gisele.secco@ufsm.br. ORCID 0000-0003-3290-746X.

fense of the reading key used here in terms of its potential for the research in philosophy of mathematics.

Keywords: Symbolic Knowledge. Leibniz. Wittgenstein. Teaching Philosophy. Philosophy of Mathematical Practice.

Abertura

Fato por vezes silenciado, as relações entre práticas filosóficas e matemáticas são atávicas nas mais distintas culturas humanas. Fato bastante conhecido, reflexões sobre as especificidades, os objetos e a aplicabilidade do conhecimento matemático cruzam a inteira história da filosofia ocidental. Assim, o processo de preparar um curso introdutório – visando a familiarização de estudantes de graduação com este corpo ancestral de conhecimentos que hoje denominamos filosofia da matemática – demanda não somente um exercício de seleção de temas, problemas, autores e textos, como também uma revisão das narrativas que encontramos na literatura de apoio, em especial a disponível em língua portuguesa.

Dos três livros introdutórios mais recentemente publicados em nossa língua, a saber, *Introdução à filosofia matemática* de Bertrand Russell (2007),[2] *Filosofias da matemática* de Jairo Da Silva (2007) e *Filosofia da matemática* de Stuart Shapiro (2017), o primeiro foi escrito durante a evolução da famigerada crise nos fundamentos da matemática – por sua vez tematizada com mais ou menos robustez nos outros dois. A diferença de ordem histórica entre os textos, que implica diferenças de ordem metodológica e conceitual entre o primeiro e os outros, é relevante para as decisões programáticas do docente: por um lado, eleger a obra de Russell pareceria já implicar um posicionamento, tomada mesma de partido, na celeuma.[3] Por outro, apresentar tal amostra da prosa do mentor de Wittgenstein a nossos estudantes ocasionaria um contato positivo com um pensamento filosófico em plena forma, elaborado com a ge-

[2] Apesar de existirem traduções mais antigas, aponto somente esta.

[3] Sobre a tomada de partido por parte de professores, ainda que se trate de partidos políticos e não filosóficos, vale a pena conferir o que tem a dizer o próprio Russell em "Funções de um professor" (2000). Sobre o tema da neutralidade do professor em sala de aula – não somente de filosofia – sugiro particularmente o segundo capítulo de *Ensino de filosofia e currículo* (Rocha 2015) e o mais recente *Escola Partida: ética e política na sala de aula* (Rocha 2020).

nuína intenção de ser "essencialmente uma 'Introdução'" (RUSSELL 2007: 15) – e, vale notar, escrito sob ímpar condição.[4] Fosse como fosse, a eleição do livro de Russell no hipotético curso introdutório exigiria do docente uma apresentação do cenário no qual ele é apenas um dos elementos, somente uma das posições em ação na querela dos fundamentos. Além disso, seria pertinente lembrar que não se tratava, para Russell, de uma introdução à filosofia *da* matemática, mas de uma *filosofia matemática* (*mathematical philosophy*), ou seja, de parte daquilo que ele considerava possível chamar matemática.[5]

No que tange às conveniências didáticas que se podem obter com os demais dois livros, deve-se logo ponderar que entre eles as diferenças são outras, além da que preservam com o de Russell: ambos foram publicados recentemente, por autores de diferentes nacionalidades e envergaduras, visando oferecer um panorama de autores e problemas que abarcam mais do que as questões abordadas por Russell. Em termos de estrutura, por exemplo, o livro de Da Silva se organiza em sequência histórica: um capítulo sobre Platão e Aristóteles; um sobre Leibniz e Kant; um sobre Frege e o logicismo (com duas econômicas páginas dedicadas a Russell); um sobre o construtivismo e um último sobre o formalismo. Já a composição do livro de Shapiro articula (entre suas quatro partes e seus capítulos encadernados em quase quinhentas páginas), ao menos três grandes eixos: um metodológico, um histórico e um conceitual. Valeria notar outras diferenças, como aquela marcada pela robustez do elenco selecionado nas referências bibliográficas, ou ainda outra, relativa ao grau de aprofundamento e à sofisticação argumentativa de cada autor. Um dos aspectos em comum entre os dois livros, entretanto, me interessa. Trata-se do modo como reconstroem as discussões filosóficas constitutivas da conjuntura de crise dos fundamentos, pelo qual se constata uma ilustre ausência: a da posição do jovem Wittgenstein.

Razoavelmente, penso, é de se esperar que o docente que deseje ou ao qual seja incumbida a tarefa de lecionar um curso de introdução à filosofia da matemática seja ou bem um especialista na área ou alguém muito bem (in)for-

[4] Na prisão, para onde Russell fora levado em 1918 em virtude de declarações que supostamente insultavam aliados de guerra do Reino Unido, conforme explica Slater na apresentação da edição brasileira, traduzida por Maria Borges.

[5] Agradeço a Rodrigo Sabadin Ferreira por chamar a atenção para este ponto.

mado, quer dizer, alguém capaz de oferecer a seus estudantes a oportunidade de desenvolver suas habilidades filosóficas com base no que de melhor, de mais atual e relevante se sabe, pesquisa e publica (seja em nossa língua materna ou em línguas estrangeiras) na área a ser ensinada. Ora, considerando-se o fato de que após ter passado por um período de abandono e maus-tratos editoriais – que poderiam bem explicar o pouco-caso, as leituras caricaturais, e até mesmo algumas injustiças interpretativas que sofreu e ainda sofre –, a filosofia da matemática de Wittgenstein passa por um processo de recuperação bastante bem sucedido desde inícios dos anos 1990, quando da publicação do seminal *Wittgenstein's Philosophy of Mathematics* (Frascolla 1994);[6] considerando-se ainda que pesquisadores brasileiros estão na linha de frente dos negócios wittgensteinianos no cenário mundial, tendo promovido eventos mais ou menos inclusivos, publicado inúmeros artigos, volumes de revistas acadêmicas e livros (em português e em línguas estrangeiras) nos quais se apresentam, analisam e desenvolvem reflexões acerca da matemática do filósofo austríaco; considerando que Wittgenstein não somente não ignorava mas se posicionava constantemente (e idiossincraticamente) acerca das questões mais prementes que se debatiam entre os representantes das "três grandes" filosofias da matemática engendradas no início do século XX; considerando por fim, que isso que designamos de filosofia da matemática de Wittgenstein tem bastante a oferecer a docentes-pesquisadores e a nossos estudantes, tanto conceitual como metodologicamente – é dizer, em potencial filosófico – quer me parecer que são escassas as razões para preparar um curso de introdução à filosofia da matemática mantendo a solene lacuna observada nos dois livros há pouco mencionados.

Assim, meu *superobjetivo*[7] consiste em incluir a estranha filosofia wittgensteiniana da matemática nas narrativas sobre a *Grundlagenkrise*. Meu objetivo aqui, entretanto, é o de mostrar como, por sincopadas que sejam, as

[6] Seguido de outros trabalhos relevante do próprio Frascolla (1997, 2017) e, para mencionar alguns, o livro de Mathieu Marion (1998), diversos trabalhos de Juliet Floyd, (2002, 2007a), Felix Mühlhölzer (2005). Em português, além de alguns certeiros parágrafos do ensaio introdutório à sua tradução do *Tractatus* (Santos 2000), temos Santos (2008). O trabalho mais importante que é a premiada tese de doutorado de Anderson Nakano (2015). Notável antecipação da qualidade das abordagens pós-Frascolla, deve-se lembrar de Wrigley (1977).

[7] Cf. D'Agostini (2018).

observações tractarianas sobre a aritmética merecem ser consideradas em suas especificidades, como saliências que permitem melhor posicionar o filósofo quando jovem no contexto logicista. Para tanto, além desta abertura, compus este texto em três partes: na primeira, ofereço o contexto *interno aos estudos wittgensteinianos* a partir do qual, via trabalhos de Sören Stenlund (2012, 2013, 2014, 2015), valido a estratégia de traçar conexões entre Leibniz e o jovem Wittgenstein *no que diz respeito à matemática*, apresentando também a noção de conhecimento simbólico leibniziano por meio de ideias de Oscar Miguel Esquisabel (2012) e Abel Lassalle Casanave (2012). Na segunda parte, desdobrando o que fizemos em trabalhos conjuntos anteriores, (Secco & Noguez 2017), forneço uma comparação entre os principais traços da noção leibniziana de conhecimento simbólico e algumas passagens tractarianas sobre a aritmética. Esta chave de leitura permite a um só tempo (i) projetar nova luz sobre as especificidades da definição tractariana de número em comparação com as de seu amigo Russell e de seu admirado Frege; (ii) fazer despontar a compreensão da natureza do conhecimento matemático como conhecimento simbólico ou formal do jovem Wittgenstein; (iii) elencar algumas razões para a alegação de que Wittgenstein pode ser considerado o filósofo da prática matemática *avant la lettre*. O trabalho se encerra com um apanhado, um retorno à reflexão inicial sobre os vínculos entre pesquisa e ensino, e uma defesa da chave de leitura aqui utilizada em termos de seu potencial para o desenvolvimento de estudos wittgensteinianos internos e externos à comunidade de especialistas.

1. Matemática e conhecimento simbólico: conexões

1.1. A abordagem de Sören Stenlund

Embora sejam por vezes reconhecidas, as semelhanças entre o tratado de juventude de Wittgenstein e a extraordinária obra de Leibniz,[8] no que tange às suas perspectivas acerca da matemática, não tem sido objeto de muitos estudos por parte dos especialistas na obra do vienense. Significativa exceção se encontra na série de traba-

[8] Sobre este ponto ver Hydé Ishiguro (1972) e Juliet Floyd (2007b).

lhos em que Sören Stenlund (2012, 2013, 2014, 2015) se dedicou a pensar as especificidades das observações de Wittgenstein sobre a matemática através da vinculação com o que chamou de "matemática simbólica" ou "matemática moderna":

> The notion of symbolic mathematics has its roots in the invention of algebra and the algebraic methods in the 17th century. F. Vieta (1540–1603), who made decisive contributions to this development, uses the word 'symbol' in the sense that is relevant here. Symbolic mathematics is to be contrasted with mathematics as an ontological science, for instance as the science of quantity and magnitude, which was the prevailing view in ancient Greek mathematics and in particular in the renaissance version of the Aristotelian and Euclidian heritage. (STENLUND, 2013, p. 8)

Essa caracterização é acompanhada de uma série de considerações histórico-filosóficas, nas quais procura mostrar que o advento de técnicas e simbolismos algébricos foi tão poderoso a ponto de ter possibilitado a Descartes a criação de sua geometria, e a Leibniz a invenção do cálculo integral. Com base nisso, Stenlund afirma não somente a permanência da concepção ontológica de matemática no interior da lógica formal, moderna lógica matemática, como também defende a tese geral de que existe, desde o *Tractatus*, uma forte comunidade de perspectiva entre Wittgenstein e a tradição algébrica, que nomeia matemática simbólica.

Dessa tese geral sai uma segunda, específica, de acordo com a qual sua tese geral é o pano de fundo mais adequado para compreender um aspecto fulcral do que chamamos de filosofia da matemática de Wittgenstein, sua postura frente a certos exageros de alguns discursos, prosas filosóficas acerca da lógica, da matemática, e da lógica matemática. Os nomes que tem sido dados para uma tal postura, ainda mais se levarmos em conta as flutuações típicas da fase intermediária do pensamento do austríaco, são muitos: antiplatonismo, antirreferencialismo, antiproposicionalismo, quase-formalismo, antirrealismo, realismo à moda... Dentre tais possíveis títulos, Stenlund elege o aspecto a partir do qual o contraste entre Wittgenstein e os demais filósofos da matemática de seu tempo é radical, o de antifundacionalista.[9] Uma terceira tese de

[9] Para uma abordagem do antirrealismo e do antifundacionalismo de Wittgenstein em filosofia da matemática, na qual se mostram fortes linhas de conexão entre a sua e a filosofia da matemática de Henri Poincaré, com extremo proveito se pode ler a tese de doutorado de Camila Jourdan (2009).

Stenlund deve ainda ser destacada, a saber, a de que a via de conexão com a perspectiva da matemática simbólica tem um papel hermenêutico unificador: "a concepção simbólica da matemática nos oferece uma perspectiva a partir da qual a unidade da filosofia da matemática de Wittgenstein torna-se patente". (2013, p. 14) Dessa tese, entretanto, não tratarei aqui.

Sobre o *Tractatus*, Stenlund considera que em grande parte o livro é associável à tradição da matemática simbólica, posto que, em primeiro lugar, a ideia crucial de sintaxe lógica está diretamente relacionada com a atividade de manipulação simbólica, com o uso regrado de signos, como expresso claramente nas passagens 3.326 e 3.327. Em segundo lugar, nada menos do que o pensamento fundamental do livro (seu *Grundgedanke*), é o de que as constantes lógicas nada representam ou substituem (*nicht vertreten*), tal como o fazem nomes com relação a objetos (4.0312). Ademais, Wittgenstein sentencia no início do segundo grupo de passagens sobre a matemática (6.2) que as equações, as "proposições da matemática", são *pseudoproposições* (*Scheinsätze*) – ao contrário do que pensava Frege, elas não expressam pensamento, não veiculam informação que possa ser verdadeira ou falsa. Ao contrário, enfatiza Stenlund, as equações são expressões que autorizam ações ou operações (mas não no sentido das operações de verdade, cruciais para a teoria da proposição desenvolvida no livro). A seguinte passagem resume bem o espírito da conexão proposta por Stenlund entre o *Tractatus* e a matemática simbólica:

> It [logical syntax] has to do not only with the external combinatorial forms of the signs as finite sequences, but with *the forms of use of signs*, i.e. *the forms of use that constitute the symbols*. The logical syntax of the *Tractatus* needs no supplementation with something like a 'semantics', because it is *already* concerned with the sense of signs: to get clear about the forms of use of signs is to get clear about their sense (or logical content). (STENLUND, 2013, p. 25)

O trecho ocorre no contexto em que Stenlund está defendendo a interpretação de acordo com a qual a concepção de sintaxe lógica do jovem Wittgenstein difere significativamente daquela herdada das concepções semânticas sobre a linguagem. Enquanto nelas a sintaxe é tomada mais como um *sistema estático*, para Wittgenstein se trata antes de tudo de algo com o qual se opera, uma *atividade*. O filósofo sueco quer sublinhar a ênfase de Wittgenstein no as-

pecto operacional do simbolismo como decisiva para a economia interna do *Tractatus* (3.326 reza que é preciso atentar para o *uso significativo* dos sinais para que neles sejam reconhecidos [*erkennen*] símbolos; enquanto 3.327 determina que signos só podem fixar formas lógicas se forem considerados em conjunto com seu *emprego lógico-sintático*). Além disso, como indicado acima, Stenlund sustenta que a operacionalidade, o caráter acional do simbolismo do qual a sintaxe lógica é uma expressão, é uma das linhas de continuidade mais prementes no pensamento de Wittgenstein, mesmo ao longo do camaleônico período intermediário (entre 1929, quando retorna oficialmente às práticas filosóficas, e meados de 1935).

É relevante notar também que inicialmente os argumentos de Stenlund (2012, 2013) para mostrar os vínculos entre Wittgenstein e a matemática simbólica, capazes de explicar a singularidade da perspectiva do filósofo quando jovem, cursavam um caminho via Spengler e Hertz:

> I don't want to say that Wittgenstein *was inspired* by 17th-century mathematicians. He had surely become aware of the deep difference between ancient Greek mathematical thinking and modern mathematics through his reading of Spengler in the early 1930's. But the symbolic point of view is present already in the *Tractarian* conception of arithmetic and logic. I think that the main source of inspiration for Wittgenstein were certain ideas issuing from the struggle for rigor in mathematics and theoretical physics in the latter half of the 19th century, in which the symbolic point of view was prominent. *He was particularly influenced* in this respect by Heinrich Hertz. Hertz' work on the mathematics of classical mechanics, in which Hertz showed how to deal with conceptual problems connected, for instance, with the notion of force of classical mechanics, was an important and influential contribution to the symbolic point of view. (STENLUND 2013, pp.21-22, meus itálicos)[10]

Já em trabalhos posteriores, embora ainda considere a filosofia tractariana da matemática como representante do ponto de vista simbólico e mantenha que não se trata de uma influência direta, Stenlund (2014, 2015) passa a caracterizar o pensamento do jovem Wittgenstein como legítimo emissário do modo leibniziano de conceber o pensamento simbólico *strictu sensu*. Recor-

[10] O vínculo ente as concepções de Wittgenstein e Hertz por meio da teoria leibniziana da expressão simbólica, em especial o modo como Wittgenstein entende o papel do simbolismo na física, pode ser investigado em Massimo Ferrari e Ion-Olimpiu Stamatescu (2002).

rendo a alguns elementos da noção leibniziana de conhecimento simbólico tal como analisados em um ensaio de Oscar Miguel Esquisabel (2012), o filósofo sueco sublinha com ainda mais força a importância dada por Wittgenstein ao aspecto operacional ou acional da matemática. Ocorre, entretanto, que ao se apropriar dos traços com os quais Esquisabel delineia a noção de conhecimento simbólico em Leibniz, Stenlund reconhece apenas *en passant* a outra feição desse conceito *bifrons*: seu aspecto *ectético*[11] ou de expressão visual – a meu ver vital para uma adequada compreensão das observações de Wittgenstein sobre a aritmética em seu tratado.[12] Passo, então, a um resumo das principais notas características da noção leibniziana de conhecimento simbólico de acordo com Esquisabel.

1.2. Conhecimento simbólico leibniziano: traços

Leibniz introduz a noção de *pensamento simbólico* – nomeada de diversos modos em muitos de seus escritos filosóficos –[13] já em sua *Dissertatio de Arte Combinatoria* (1666). O contexto mais amplo no qual é engendrada, vale notar, é a distinção cartesiana entre conhecimento intuitivo (não mediado) e conhecimento simbólico (mediado por signos). Esquisabel explica que,

[11] O termo "ectético", explica Esquisabel, vem do grego *ékthesis* e, literalmente, significa exposição ou apresentação. Em sentido técnico é utilizado para se referir a tipos de provas ou demonstrações, tanto em contexto aristotélico (uma prova na qual se introduz um termo singular, instância ou exemplificação, na inferência) como em contexto euclidiano (o passo de uma demonstração no qual se apresentam ou instanciam as informações geométricas contidas na proposição ou problema a ser provado ou resolvido). Tal como utilizado por Leibniz, *ékthesis* tem relação com o modo como Jungius o aplica em sua discussão sobre demonstrações silogísticas que podem ser modificadas de tal modo que a forma lógica de uma proposição possa ser representada: "From the fact that the syntax of formulae makes a form or structure ostensible it follows that formulae have an ambivalent trait, since on the one hand they are suitable for mechanical computation and on the other they are *displaying something* that goes beyond themselves. Furthermore, Leibniz sees in this 'expositive' capability of formulae the grounds for their performance in invention or discovery, because they make us discover relations or connections that otherwise would be hidden to us, a feature that we could call 'expressivity' in speaking of formulae. For this reason, *formulae share some features with pictures and diagrams*, which also play a role as perceptible guides for thinking, although not necessarily in a mechanical fashion" (ESQUISABEL, 2012, p. 29, meus itálicos).

[12] A articulação entre operar e exibir, como duas faces de todo o simbolismo, pode bem ser estendida para as observações de Wittgenstein sobre a lógica – o que ela faz e o que ela mostra. Tal ampliação, entretanto, não cabe no espaço destas notas.

[13] "Such denominations are "blind thought" (*cogitatio caeca*, *pensée sourde* or *vuide*), "blind knowledge" (*cognitio caeca*), "suppositive notion" (*cognitio suppositiva*, *connaissance suppositive*) and even "suppositive notion" (*notio suppositiva*) and "blind concept" (*conceptus caecus*)." (ESQUISABEL, 2012, p. 2)

na obra inaugural de Leibniz, a metáfora do *pensamento cego* ocorre na caracterização de número. A seguinte passagem é citada para mostrar como Leibniz entende a "performance simbólica do número":

> For example, we often grasp a number, however large, all at once in a kind of *blind thought*, namely, when we read ciphers on paper which not even the age of Methusela would suffice to count explicitly. (A VI 1, 170 [Loemker 76], *apud.* ESQUISABEL, 2012, p. 12)

O que aqui merece atenção é a ideia de simultaneidade de percepção que um só signo (digamos, 10^{100}) favorece quando deputa ou está por um todo bastante largo, impossível de ser apreendido nas unidades que o compõem. De acordo com Esquisabel essa ideia faz emergir outra, que antecipa um aspecto fundamental das concepções leibnizianas sobre a representação simbólica: de que "o número é um tipo de 'figura universal' por meio da qual *todas as coisas*, seja qual for sua natureza, podem ser *representadas e trabalhadas*. (ESQUISABEL, 2010, p. 12, meus itálicos). Importa notar que o número não é entendido como "figura universal" porque a linguagem da aritmética seja capaz de representar, ou designar, do mesmo modo como nomes ou proposições representam. Para Leibniz, o número é uma figura universal porque a manipulação do simbolismo aritmético organiza-se de modo a sub-rogar aspectos estruturais de qualquer representação conceitual – algo que é também característico do modo algébrico de pensar, afinal a grande inspiração do projeto da *Característica universal* leibniziana.

Da utilização de números, constantes individuais, variáveis, sinais de operação e de igualdade, etc., diz-se sub-rogatória no sentido de que toda manipulação de símbolos no interior do sistema no qual que se faça corresponder a conceitos ditos analisáveis pode ser reduzida a uma manipulação dos símbolos cuja operação ocorra em quaisquer níveis ulteriores de análise lógica desse conceito. Na alvorada da tradição do conhecimento simbólico, a noção de número aparece como consideração abstrata – e aplicável variavelmente – da relação entre todos e partes expressa em qualquer juízo. Assim, fica claro que a substituição da linguagem na qual um juízo qualquer é formulado pelo arcabouço simbólico aritmético (ou algébrico) não altera nem acrescenta conteúdo cognitivo (proposicional), mas apenas *o exibe* conforme às relações estruturais

relevantes para associá-lo ao que é expresso por outras formulações da linguagem. A primeira e bastante geral caracterização da noção de pensamento cego, já se pressente, está conectada ao conceito aritmético de número, cuja definição marcará uma diferença crucial do *Tractatus* para com as concepções de Frege e Russell. Mas não nos antecipemos. Ofereço a seguir, listados, os principais aspectos ou funções do pensamento simbólico cego tal como apresentados por Esquisabel e apropriados por Stenlund:[14]

> **(a) função sub-rogatória** ou representacional, pela qual signos artificiais (como signos algébricos ou aritméticos) substituem ou estão por alguma *ideia*;
>
> **(b) função psicotécnica**, de abreviação e simplificação de processos de pensamento, que permite aprimorar nossas capacidades de raciocínio e, portanto, de obtenção de conhecimento seguro. Sem precisar considerar a compreensão dos "significados materiais" dos signos, alivia-se a memória de quem raciocina;
>
> **(c) função calculatória**, computacional ou de simbolismo operativo, pois pensamento simbólico cego requer certas "estruturas tomadas como sistemas de objetos físicos sujeitos a operações de construção e transformação de acordo com regras", preservando uma ordem estrutural ou sintática – donde a ideia leibniziana de um *filum mechanicum meditandi* que, a seu turno, está diretamente associado o aspecto seguinte;
>
> **(d) função *ectética*, ou de exibição**, associada à demanda pela apreensão de formas ou estruturas exibidas nas configurações simbólicas elas mesmas, de modo que o pensamento opere "por meio de uma espécie de raciocínio visual." Este é um aspecto de importância central para a compreensão da natureza formal ou estrutural do pensamento simbólico, e mesmo para sua função sub-rogatória (isso porque "a estrutura do objeto ou estado de coisas é projetada na sintaxe da expressão simbólica");

[14] Trechos desta subseção foram emprestados, com ligeiras modificações de meu já referido trabalho com Secco & Noguez (2017).

> **(e) função instrumental** que se manifesta pelo uso de regras operacionais de construção e transformação de signos, nos quais as próprias expressões guiam e permitem verificar as inferências ou cálculos em jogo (Esquisabel fala, a esse respeito de verificabilidade *ad oculos*); o pensamento assim se torna como que imune às compreensões confusas, típicas do comércio comum da fala, tributário do pensamento verbal;
>
> **(f) função que possibilita conhecimento de estruturas ou formas**: uma vez que as inferências levadas a cabo neste modo de pensar abstraem os significados materiais dos termos envolvidos (as ideias ou conceitos que eles sub-rogam), o que resta desta abstração – as características formais dos sistemas simbólicos nos quais se opera – pode ser investigado em si mesmo, de modo igualmente abstrato e formal. Sistemas simbólicos, afinal, podem expressar-se mutuamente – conforme se abstraiam as relações formais de cada um, invariantes sob as transformações projetivas ou mapeamentos entre sistemas simbólicos. Ao fim e ao cabo, anote-se, estas relações desempenham um papel determinante para nada menos do que a caracterização leibniziana de verdade (já que seriam sua "base"). Alcança-se, desse modo, um tipo especial de conhecimento: sintático e estrutural.

Em resumo, as fórmulas e concatenações de fórmulas de cunhagem algébrica, que inspiram a ideia deste pensamento *sui generis*, sub-rogam, abreviam e simplificam processos de raciocínio, são mecanicamente manipuláveis, exibem ou expressam relações estruturais, permitem a verificação *ante oculos* de tais relações que podem existir entre diferentes estruturas simbólicas, e, assim, engendram conhecimento específico, conhecimento de tipo formal. Vale notar que para Leibniz não se trata aqui de uma modalidade de conhecimento *a mais*, mas sim de um gênero de conhecimento capaz de proporcionar o mais alto grau de certeza possível ao intelecto humano – o que é metaforicamente expresso na ideia da possibilidade de um teste *ante oculos* da verdade por meio de um *filum meditandi*.

O projeto leibniziano de uma ciência geral de estruturas formais instanciadas em qualquer domínio particular de conhecimento – a *Característica Geral* – é o ápice do conceito leibniziano de conhecimento simbólico cego. O objetivo da *Característica Geral* consiste em investigar as leis de composição e transformação de formas "vazias", fornecidas pelas propriedades das fórmulas que exibem formas gerais de relações, conexões e operações que não possuem interpretação determinada. Pode-se compreender o projeto como o delineamento de uma *ciência ou teoria ectética abstrata de nível superior*, que lida com estruturas teóricas. Esta ciência possibilita *conhecimento simbólico de invariantes estruturais*, "das condições de possibilidade do conhecimento simbólico como tal", revelando-se seus principais traços, ao modo de *Janus bifrons*: uma de suas feições é a de uma sintaxe operatória e a outra de expressar-se visualmente.[15] Embora seja caracterizado pela capacitista metáfora do *pensamento cego*, o conhecimento simbólico leibniziano depende essencialmente de certo tipo de visão.[16] Antes de indicar as reverberações desta imagem Januária no tratado de Wittgenstein devo ainda fazer um par de anotações, como complemento à abordagem de Esquisabel, complemento decisivo para o movimento que se segue na próxima seção. Trata-se da apresentação do conhecimento simbólico oferecida por Abel Lassalle Casanave (2012, 2016).[17] Lembrando que a distinção cartesiana entre conhecimento intuitivo e simbólico é um elemento contextual importante para a compreensão da noção engendrada por Leibniz, a compreensão do filósofo portenho contém três variações:

> **CS_{SI}** – conhecimento simbólico como *sucedâneo do conhecimento intuitivo*, quando "o simbolismo teria sempre uma contrapartida intuitiva de forma tal que alcançamos por seu intermédio um conhecimento que

[15] Sobre as inúmeras dimensões do programa de Leibniz para uma *Característica geral* ver Esquisabel (2002).

[16] Não parece desacertado afirmar que se trata de um *ver-como*. Contemporaneamente, o tópico da visualização nas ciências formais é bastante discutido, em especial por filiados ao vago rótulo de filosofia da prática matemática, que abordarei ao final destas notas. Para uma excelente coletânea em língua portuguesa, da qual destaco – dados os fins deste texto – os capítulos de Paolo Mancosu, Abel Lassalle Casanave, Camila Jourdan e Luiz Carlos Pereira, ver Lassale Casanave & Sautter (2012).

[17] Respectivamente, em um texto sobre o conhecimento simbólico em Kant e outro sobre conhecimento simbólico e aritmética em Hilbert.

> também poderíamos alcançar intuitivamente" (LASSALLE CASANAVE 2016, p. 227);
>
> **CS_{EI} –** conhecimento simbólico como *extensão instrumental do conhecimento intuitivo*, quando "são introduzidos por via simbólica conceitos e métodos sem contrapartida intuitiva, cuja finalidade é alcançar conhecimento que também poderíamos alcançar por procedimentos intuitivos, mas cuja execução eventualmente seria muito complexa" (*ibid*, p. 228);
>
> **CS_{CF} –** *conhecimento simbólico como conhecimento de estruturas, formas ou relações*, "em cujo caso falamos de conhecimento simbólico como conhecimento formal." (*loc.cit.*).

Resumindo, pode-se conceber que a **CS_{SI}** de Lassalle Casanave corresponde primordialmente a função **(a) sub-rogatória** de Esquisabel – de acordo com a qual o conhecimento simbólico depende de signos que substituem ou designam ideias; **CS_{EI}** enfatiza a função **(e) instrumental**, associada à **(d) função ectética**, posto que as operações (inferências, cálculos e/ou demonstrações) são norteadas pelo simbolismo ele mesmo, a correção da aplicabilidade de suas regras se verificando *ante óculos* etc.; **CS_{CF} –** identifica-se à função **(f)**, que chamaria de **função cognitivo-estrutural**, fosse eu autorizada a rebatizá-la.

Imagino que a esta altura a leitora frequente do *Tractatus* já tenha antecipado diversas comparações que se podem mobilizar no intuito de evidenciar elementos leibnizianos nas passagens sobre a matemática. Supondo, entretanto, que nem toda leitora destas notas tenha já, por qualquer uma de suas entradas, se embrenhado no labirinto, delineio possíveis cursos na próxima seção.

Para finalizar, retomo a Stenlund. Como indiquei acima, ainda que considere quase todos os aspectos **(a)–(f)** da noção de conhecimento simbólico na caracterização das reflexões de Wittgenstein acerca da matemática, Stenlund enfatiza significativamente o aspecto operacional, em detrimento do aspecto expressivo ou *ectético* dos simbolismos algébricos que inspiram Leib-

niz e ecoam nas reflexões Wittgenstein. Essa ênfase se explica, ao menos parcialmente, pelo fato de que o filósofo sueco estava mais interessado em sustentar a tese da unidade do pensamento de Wittgenstein sobre a matemática, que a chave leibniziana de leitura favorece, enfatizando a importância da ideia de uso dos signos, do que explorar suas nuances no contexto do *Tractatus.* Enquanto Stenlund interessa-se pelos desenvolvimentos de ideias tractarianas no camaleônico período intermediário, o meu é um recorte mais modesto.

2. Legados leibnizianos no *Tractatus*: o conceito de número e a ideia de conhecimento matemático

A metáfora mais utilizada em alusões, referências, análises especializadas e textos de introduções à leitura do livro de Wittgenstein é aquela que próprio autor forjou entre parêntesis na penúltima passagem [6.54].[18] Penso, entretanto, que a imagem do *Tractatus* como um labirinto com múltiplas entradas é mais adequada, especialmente se se considerar minha breve discussão inicial sobre como ensinar introdução à filosofia da matemática *com Wittgenstein*. Diferentemente da metáfora da escada, que exige uma leitura em progressão sequencial (cada uma das sete passagens do livro sendo um grande degrau – não se pode subir uma escada começando por qualquer degrau) – a imagem do labirinto de múltiplos inícios oferece uma liberdade que, para fins didáticos, é preciosa. Não que não seja preciso fornecer aos estudantes exposições e leituras que os permitam *saber a quantas andamos* (*knowing our/your way about*) quando enfrentamos a argumentação sincopada que encaramos ao ler o *Tractatus*. Antes de tudo, por certo, é preciso que já os tenhamos familiarizado com o contexto, quer dizer, com as filosofias da aritmética contra as quais Wittgenstein está a seu modo se rebelando – bastariam boas leituras prévias das partes mais relevantes dos *Fundamentos da Aritmética* de Frege e dos três primeiros da *Introdução* de Russell, pelos quais se podem ensinar as diferenças conceituais nuançadas que subjazem suas definições de número, mas também aquilo que, tendo em comum, será objeto da crítica de Wittgenstein: o fato de que são definições dependentes da noção de classe, de uma analogia

[18] Para algumas "instruções sobre como subir a escada" ver Engelmann 2018a e 2018b.

entre o modo de representar de nomes e de símbolos lógicos e matemáticos e, claro, de serem definições que, cada uma a seu modo, fazem dos números objetos.

Em momento oportuno (o que não quer dizer de início, nem na ordem em que aparecem no livro, afinal a lógica dos conteúdos não corresponde à lógica das aprendizagens),[19] é fundamental que os estudantes façam ideia da teoria geral da proposição, a seu turno dependente das ideias de fato e de pensamento, das noções de operação, de função, de verdade *et cetera*;[20] mais importante, é preciso alguma nitidez sobre o sentido do *Grundgedanke*, exposto em 4.0312 no contexto de construção da teoria da proposição – o de que, contra Frege e Russell, "as constantes lógicas não substituem", quer dizer, não representam ou designam como os nomes fazem com os objetos. Devem os estudantes estar cientes, por certo, da ideia de *forma geral da operação*, que antecede ao primeiro grupo de "passagens aritméticas" – formado por meros cinco aforismos [6.02 – 6.31]. Neles, encontramos: uma definição de número (inteiro) [6.02, 6.021]; algumas considerações sobre a "forma geral" e o conceito de número inteiro [6.022, 6.03]; derivada do que precede, uma acusação de superfluidade direcionada à teoria de classes na matemática [6.031]; a alegação de que esta desnecessidade vincula-se ao tipo de generalidade, não causal, que a matemática demanda.

A definição de número é dada, como indiquei, a partir do que Wittgenstein chama de "forma geral da operação", oferecida em 6.01. Uma operação caracteriza-se como a ação de engendrar uma proposição a partir de outras proposições ou, como explica Marion "mais precisamente como aquilo por meio do que uma forma de proposições é engendrada a partir de outra forma de proposições (5.23)" (2012, p. 90). A forma da passagem de uma proposição à outra se dá por meio da aplicação de um operador, o chamado operador *N*.[21] Além da forma geral da operação, a definição de Wittgenstein

[19] Esta lembrança é feita em Rocha (2015), o mesmo autor que intuiu a didática como a *arte da graça*.

[20] Há, em português, dois livros que, combinados, podem ser uma boa introdução geral à leitura do *Tractatus*: Marion (2012) e Marques (2005). Uma resenha valiosa do primeiro foi escrita por Silva (2015).

[21] Sobre o qual recomendo fortemente a leitura da dissertação de Rodrigo Sabadin Ferreira (2017).

depende de regras notacionais estipuladas na primeira parte de 6.02. Esta notação pode ser apresentada através da interpretação de Frascolla:

> ***x*** exibe a forma de uma expressão que ainda não foi gerada pela aplicação de uma operação.
>
> **Ω** é a operação variável, o símbolo para o conceito de operação.
>
> **'** é forma do resultado da aplicação de uma operação a uma base dada; resultado de uma operação em geral.

Assim, diz Wittgenstein, em 6.02, "chegamos ao conceito de número:

> defino
> $x = \Omega^{0'} x$ Def.
> $\Omega' \Omega^{v'} x = \Omega^{v+1'} x$ Def."
> (WITTGENSTEIN, 2001, p. 249)

A partir destas definições, e por meio da aplicação sucessiva das regras sobre uma mesma base arbitrária, ou seja, por indução – ou *definição recursiva*, em que se define um objeto em termos de si próprio – gera-se a série (formal) ordenada dos números naturais. Assim, produz-se a seguinte sequência:

$$
\begin{array}{rcll}
x = \Omega^{0'} x & \rightarrow & 0 & = 0 \quad Def. \\
\Omega' x = \Omega^{0+1'} x & \rightarrow & 0+1 & = 1 \quad Def. \\
\Omega'\Omega' x = \Omega^{0+1+1'} x & \rightarrow & 0+1+1 & = 2 \quad Def. \\
\Omega'\Omega'\Omega' x = \Omega^{0+1+1+1'} x & \rightarrow & 0+1+1+1 & = 3 \quad Def. \\
\Omega'\Omega'\Omega'\Omega' x = \Omega^{0+1+1+1+1'} x & \rightarrow & 0+1+1+1+1 & = 4 \quad Def. \\
\Omega'\Omega'\Omega'\Omega'\Omega' x = \Omega^{0+1+1+1+1+1'} x & \rightarrow & 0+1+1+1+1+1 & = 5 \quad Def. \\
 & & \vdots &
\end{array}
$$

Figura 1

Reconstrução da definição tractariana de número por Renato Reis Leme (2018)

Wittgenstein completa sua definição determinando, em 6.021, que "O número é o expoente de uma operação." Com isso – quer dizer, tomando como primitiva a noção de operação e a ideia de expressão, ou do mostrar-se,[22] o filósofo se contrapõe inteiramente aos seus antecessores logicistas, em especial a Frege, cuja concepção de número era objetual. Para Wittgenstein, números *expressam*, como traços, a posição que ocupa, em uma série formal, aquilo (a base) sobre o que foi aplicada uma operação. Como nota Noguez, "operações se mostram em suas aplicações; não são, não se confundem com suas aplicações" (2018, p. 44). Embora talvez desagradasse a Wittgenstein, não seria equivocado dizer que da síntese entre sua definição de número e o *Grundgedanke* do tratado, a ideia de que números não são objetos é um corolário.[23]

A seguir, em 6.022, Wittgenstein afirma que o conceito de número "nada é senão *o que todos os números têm em comum*, a *forma geral do número*." (*loc. cit*., meus itálicos) Agora, lembremos que para Leibniz, como acima indicado, o número é uma *figura universal* porquanto os símbolos aritméticos que se manipulam por meio de regras sub-rogam *aspectos estruturais* de qualquer representação conceitual. Números, assim como constantes lógicas, para Wittgenstein nada substituem (*nicht vertreten*), não designam ao modo dos nomes, não são propriedade de um *algo*. "Número" é um conceito formal, e enquanto tal, "é já dado com um objeto que sob ele cai" ([4.12721] WITTGENSTEIN, 2001, p. 187) – neste caso, a aplicação de uma operação, que é algo *que se faz*, não que se enuncia. Didaticamente "seja o estádio do Maracanã em dia de jogo do Flamengo o objeto e a contabilização de torcedor a operação, se ambos estão precisamente definidos, então o resultado da operação nos oferecerá a cardinalidade do conjunto objeto." (LEME, 2018, p. 38). Mas talvez fosse mais adequado dizer que "o resultado da *aplicação da operação mostrará* a cardinalidade do objeto". De todo modo, destacada a in-

[22] Para esclarecimentos importantes acerca da famigerada distinção entre o que se diz e o que se mostra são inultrapassáveis tanto Floyd (2007) quando Narboux (2014).

[23] A caracterização de corolário não agradaria a Wittgenstein pelo seu tom axiomático, posto que, justamente, um dos pontos fixos de sua crítica a Frege e Russell, seu projeto logicista de fundamentação da aritmética na lógica, está associado ao modo axiomático de conceber a lógica e sua redução da aritmética, movimento cujas consequências filosóficas desagradavam a Wittgenstein – em especial como nota Marion (2012, p. 102), a consequência de que a aritmética é tão analítica quanto a lógica.

separabilidade do caráter acional ou operacional **(c)** e do caráter expressivo **(d)** da definição wittgensteiniana de número, a ressonância leibniziana da *complementaridade entre operar e exibir* torna-se clara,[24] e com ela se adquirem elementos adicionais para a compreensão das especificidades de sua noção de número como crítica às de Frege e Russell.[25]

Imagino que nesse ponto os leitores especializados, conhecedores da ênfase da literatura secundária na noção de operação como marca maior das filosofias tractarianas da lógica e da aritmética, podem estar se perguntando qual é a novidade que minha leitura traz à praia. A pergunta se justificaria ainda mais se lembrássemos que a tradição do conhecimento simbólico de corte leibniziano encontra ecos também na obra de Frege, como bem mostrou Javier Legris (2012), de modo que também nisso Frege teria influenciado Wittgenstein e que, portanto, nada há de muito relevante aqui. Ora, no que diz respeito à participação de ambos na linhagem leibniziana do conhecimento simbólico, mostramos em Secco & Noguez (2017) que as semelhanças não são tão relevantes quanto as diferenças; ainda que Frege busque, em especial na *Conceitografia*, desenvolver um sistema simbólico no qual as relações formais sejam exibidas com a maior clareza possível, a robustez de seu platonismo o distingue tanto de Leibniz quanto de Wittgenstein. Nos últimos, repito, no que diz respeito aos simbolismos lógicos e matemáticos, trata-se muito menos de considerar o aspecto sub-rogatório **(a)**, do que a aliança entre o caráter operacional e o ectético na articulação, para não dizer novamente da *complementaridade entre operar e exibir*.

O segundo grupo de passagens tractarianas sobre a matemática fornece ainda mais elementos para a consideração das diferenças entre Wittgenstein

[24] Ressonância que também encontrada em Hilbert e Bernays, como mostra Lassalle Casanave (2016, p. 229 *ssq*). Uma crítica da triangulação Leibniz-Hilbert-Wittgenstein, tematizada em Stenlund (2015), ainda está para ser elaborada. Tardo talvez em marcar o reconhecimento de que é possível também traçar um triângulo Leibniz-Russell-Wittgenstein, senão linhas de influência ou recepção direta (aqui se trata mais de um exercício de história das ideias), de linhas de conexão que ajudariam a percorrer melhor outras passagens do *Tractatus*; tratar-se-ia nesse caso de outra entrada no labirinto.

[25] É certo que existem ainda outras críticas de Wittgenstein às definições dos logicistas, que não abordarei, como a de que a definição relação de ancestralidade, fundamental para as caracterizações de Frege e Russell da sequência dos naturais, é circular – crítica, aliás, já formulada por Poincaré, como indica Marion (2012); ou a da superfluidade da noção de classe na matemática.

e seus mentores desde a perspectiva do conhecimento simbólico. Nelas, se enuncia que as proposições da matemática não exprimem pensamentos – que, em realidade, sendo equações (regras de autorização de operações aritméticas), são *pseudoproposições* [6.2, 6.21, 6.211, 6.24]; que, por meio do sinal de igualdade (uma autorização para a substituição de expressões), as equações aritméticas mostram, exibem "a lógica do mundo" [6.22, 6.23, 6.232]; que a correção das equações aritméticas "é algo a ser visto", não asserido (e nisso consiste "demonstrar": mostrar a substitutibilidade das expressões que ladeiam o sinal de identidade por meio de operações de iteração simbólica) [6.231, 6.2321, 6.2322, 6.2323]; que a intuição necessária à solução de problemas matemáticos é fornecida *pela linguagem ela mesma* – e daí resulta que calcular e realizar experimentos sejam processos conceitualmente distintos [6.233, 6.2331]; que a matemática é um método da lógica; [6.234, ecoando a primeira sentença de 6.2]. E, por fim, Wittgenstein "Demonstra" (calcula de acordo com as regras anteriormente explicitadas) que 2 x 2 = 4.[26]

Uma análise detida de todas estas ideias está fora do alcance dessas notas. O que faço aqui é considera-las em resumo e em conjunto com vistas a defender a ideia de que por mais que a letra do texto tractariano não nos autorize a falar de conhecimento matemático em sentido "genuíno" (como conhecimento proposicional), o modo como Wittgenstein se manifesta sobre as práticas matemáticas cotidianas, em especial de calcular, é suficiente como para legitimar a tese de acordo com a qual, no *Tractatus*, conhecimento matemático é conhecimento formal em estrito sentido leibniziano (**CS_{CF}** ou **(f)** acima).

Para tanto, começo destacando as passagens 6.232 e 6.2321 – esta, aliás especular àquela onde se enuncia o "fato que contém em si toda a filosofia da lógica" (6.113): o de que a verdade das proposições da lógica pode ser reconhecida (*erkennen*) no símbolo tão somente (*am Symbol allein*). A primeira parte de 6.232 se refere à ideia de Frege de que "1+1+1+1" e "(1+1) + (1+1)" têm a mesma referência (*Bedeutung*), mas diferentes sentidos (*Sinn*). E segue:

[26] Renato Reis Leme (2018) sustenta que, de fato, não se trata de uma demonstração, mas de uma computação, no sentido posteriormente determinado por Turing.

> 6.232 (...) Mas o essencial, no caso da equação, é que ela não é necessária para se mostrar *que* as duas expressões ligadas pelo sinal de igualdade têm o mesmo significado, já que isso *se pode ver nessas próprias expressões*.
> 6.2321 E que as proposições da matemática possam ser demonstradas nada quer dizer senão que sua correção é *algo a ser visto*, sem que deva o que exprimem ser comparado com os fatos quanto à sua correção. (WITTGENSTEIN, 2001, p. 263, meus itálicos)

O aspecto ectético do reconhecimento de que duas expressões equacionais expressem a mesma informação, e de que uma equação é correta só poderia ser mais diretamente exposto se Wittgenstein repetisse a formosa expressão *am Symbol allein*. Parece-me também bastante sugestivo o fato de que não se necessita (e em realidade nem sequer se possa pretender) dizer, afirmar, asserir que ambos os lados do sinal de igualdade em uma equação partilham seu *Bedeutung*. Saber usar os sinais "1" e "+", manipula-los de acordo com as regras do jogo da aritmética, é saber tudo o que é preciso para o reconhecimento da igualdade do significado de "1+1+1+1" e "(1+1) + (1+1)"; do mesmo modo, saber usar numerais e outros símbolos para operações aritméticas é saber tudo o que é preciso para que se possa legitimamente pleitear saber adicionar, subtrair, multiplicar, dividir etc.. É por isso que Wittgenstein pode afirmar em 6.2323 que equações apenas fazem assinalar um ponto de vista: o ponto de vista desde o qual consideramos as expressões que ladeiam o sinal de igualdade. Esse é o contexto no qual o tema, a questão da necessidade da intuição na solução de problemas matemáticos é introduzido, em 6.233, retomando-se sob nova roupagem o tema da verificação (reconhecimento, para o caso das proposições lógicas)[27] e da visualização (para o caso das equações) *no símbolo tão somente*.

A resposta de Wittgenstein é, mais uma vez, essencialmente leibniziana: é a linguagem simbólica ela mesma, por meio de seu aspecto calculatório (**(c)** acima), que fornece a intuição de que se necessita para resolver um problema matemático. Assim (segue a passagem 6.2331), calcular não é o mesmo que fazer experimentos justamente porque consiste em manipular

[27] Reconhecimento que constitui fundamentalmente o que é demonstrar *em* lógica (de acordo com 6.1262): dispor de um expediente mecânico (cálculo) para facilitar o *reconhecimento* (visual) de uma tautologia complicada. Demonstrar *na* lógica e demonstrar *logicamente* (*com* lógica) não é o mesmo, assim como calcular ou utilizar proposições matemáticas *na* matemática não é o mesmo que as utilizar "na vida" (6.211).

símbolos (que a seu turno *sub-rogam tão-somente conceitos e relações formais*, jamais objetos ou relações entendidas como objetos com os quais se tem alguma *acquaitance*) de acordo com regras de substituição de expressões. Em jargão tractariano, isso quer dizer que em matemática, ao calcular, estamos lidando com o âmbito da necessidade, das relações internas, relações estritamente formais. Embora também a ciência tenha um arcabouço formal de "iluminações *a priori*" (6.34), os processos de experimentação fazem uso predominante "da linguagem" (proposicional) a fim de contemplar as relações causais em jogo nos fenômenos descritos, e quando realizamos experimentos estamos na lida com o âmbito da contingência, das relações externas etc.[28]

Agora bem, é claro que se quisermos ser fiéis ao *Tractatus*, a alegação de que a matemática engendra conhecimento é descabida, pois neste livro suturado pela dialética entre o que se pode dizer e o que somente se mostra, o conhecimento se inscreve nos limites do dizível. Escovando o texto a contrapelo, minha alegação é a de que dado que se pode ver, como busquei mostrar, a presença de elementos claramente leibnizianos nas considerações do jovem Wittgenstein sobre a matemática, e que aqueles elementos, por sua vez se combinam na composição de uma noção de conhecimento de estruturas e relações formais – conhecimento formal – é válida a caracterização da filosofia tractariana da matemática, *de tout* uma filosofia da aritmética, como legatária legítima da tradição leibniziana do conhecimento simbólico.[29]

3. Encerramento: um filósofo da prática matemática extemporâneo?

A terceira e última alegação anunciada na abertura destas notas acerca da chave de leitura acima proposta diz respeito mais a uma suspeição do que a uma certeza. A suspeita tem origem no contexto das diversas investigações

[28] A dicotomia entre relações internas e externas como chave de compreensão da distinção entre asserir/dizer e exemplificar é abordada de maneira bastante clara em Jourdan (2009) e, no contexto de uma discussão sobre a autocrítica nos anos 1930, em especial o que diz respeito ao modo como criticará Frege na questão das atribuições numéricas em Velloso (2017).

[29] Para uma abordagem de outros "deslocamentos, reinvestimentos, realocações e mesmo, em alguns casos de enriquecimento, senão do ponto de vista do conteúdo, ao menos do ponto de vista de suas aplicações [dos conceitos leibnizianos em Wittgenstein]" – curiosamente sem qualquer ênfase especial na matemática – Plaud (2018).

que, por semelhança de família mais do que por definição, convencionou-se chamar de filosofia da prática matemática. Mais especificamente, quando se trata de investigações de ordem histórica – como as que encontramos na "Introdução" de Mancosu (2008) para o livro seminal desse novo *brand* de pesquisas filosóficas sobre a matemática, ou mais recentemente no trabalho de Jessica Carter (2019) – costuma-se retroceder no máximo até as investigações de Imre Lakatos em *Conjecturas e refutações*. De acordo com essa narrativa, o filósofo húngaro, sua filosofia dialética da matemática – que também chamou de quase-empirismo – teria sido um dos primeiros dissidentes das filosofias fundacionalistas da matemática que foram progressivamente se estabelecendo desde fins do século XIX até meados dos 1930.

Wittgenstein é por vezes mencionado, é verdade: como um terapeuta da linguagem cujos procedimentos podem aliviar as análises filosóficas da matemática de pesos desnecessários (Avigad 2018); como digno de desconsideração por princípio (Ferreirós 2016 – na primeira nota de rodapé de seu de outro modo importante livro); ou de modo um pouco mais simpático e operacional, como fizemos em Secco (2013) e Secco & Noguez (2017), e como faz John Mumma (2020) – sendo que nesses casos são mobilizados *insights* e outros elementos do pensamento de Wittgenstein do período maduro, quer dizer, posteriores aos anos 1939. Sobre o *Tractatus*, até hoje não me deparei com uma linha sequer em que fosse usado ou mencionado como o texto que minha suspeição indica ser: o legítimo precursor contemporâneo da postura filosófica que caracteriza, a filosofia da prática matemática. Senão vejamos.

Ora, de acordo com Mancosu os ancestrais mais relevantes das filosofias da prática matemática (o plural é por minha conta) encontram-se na *tradição dos dissidentes* (*mavericks*). Os traços gerais de suas filosofias da matemática seriam:

> **a.** anti-foundationalism, i.e., there is no certain foundation for mathematics; mathematics is a fallible activity;
> **b.** anti-logicism, i.e. mathematical logic cannot provide the tools for an adequate analysis of mathematics and its development;

> **c.** attention to mathematical practice: only detailed analysis and reconstruction of large and significant parts of mathematical practice can provide a philosophy of mathematics worth its name. (MANCOSU, 2008, p.5)

Embora reconhecendo que os autores desta tradição realizaram uma significativa extensão do domínio tradicional de estudos filosóficos sobre a matemática, eles não foram tão longe como para "redirecionar substancialmente o curso da filosofia da matemática" (*loc. cit*).

Agora, dependendo do modo como concebemos a expressão "antifundacionalismo", ela até poderia se aplicar ao caso do *Tractatus*. Não porque ali encontremos a ideia de que a matemática é uma atividade falível (segunda parte da característica **a.** de Mancosu), mas porque às críticas mais ou menos explícitas de Wittgenstein aos logicismos de Frege e Russel parecem subjazer a ideia de que a matemática não necessita de fundamentação (primeira parte da característica **a.** de Mancosu). Lembremos que, para Wittgenstein, operar com símbolos matemáticos – tanto quanto o operar com símbolos lógicos – é realizar uma prática que, famosamente, *toma conta de si mesma*, é dizer, é autônoma com relação à realidade à qual perfeitamente se aplica, embora reflita, diáfana, a lógica do mundo (6.13). Que a postura do jovem Wittgenstein seja antilogicista (característica **b.** de Mancosu), não me parece ser preciso repetir.[30] Por fim, e talvez mais contenciosamente, as observações tractarianas sobre a aritmética configuram análises de práticas matemáticas (característica **c.** de Mancosu), muito embora práticas operatórias muito simples, e detalhadas somente na medida em que permita a pré-socrática retórica do livro.

Que o tópico principal do interesse de Wittgenstein estivesse na análise de práticas triviais ou rudimentares, poderia bem indicar "um alto grau de sensibilidade conceitual por parte de sua abordagem simbólica", como já sugeriu Stenlund (2013, p. 32), que afirma ainda

[30] Embora Rodrigo Sabadin considere que eu deva sublinhar aqui que minha leitura é oposta às de intérpretes que caracterizam a filosofia da matemática do jovem Wittgenstein como um "logicismo sem classes" – Mathieu Marion – ou como um "realismo à moda" – como Luiz Henrique Lopes dos Santos).

> I would say that this is why it is so difficult in many cases to understand what he is doing, what he is up to, in his writings on the foundations of mathematics. On this level of 'trivialities', the topic of 'the nature of mathematical symbolism' is a non-issue for the mathematician. (*loc. cit.*)

Que os contemporâneos e contemporâneas filósofas da prática matemática se importem abundantemente com o tópico "a natureza do simbolismo matemático" deveria, a meu ver, servir como razão adicional para revisitar a obra de juventude de Wittgenstein, por razões históricas e conceituais. Afinal, um dos tópicos de assaz potencial está na inter-relação entre lógica, matemática e ciência da computação. Ademais, como tem indicado Juliet Floyd em diversos de seus mais recentes trabalhos, e Leme (2018) com sua ideia de programação como forma de vida, entre Wittgenstein e Turing há mais do que imagina nossa habitual filosofia.

De todo modo, e retornando à minha motivação original, terei ao menos um sucesso se leitores e leitoras engajados com filosofia da matemática, não somente como pesquisadores mas como professores-pesquisadores, considerarem a plausibilidade da ideia de que as primeiras observações de Wittgenstein sobre a aritmética merecem figurar em cursos de introdução à filosofia da matemática (e quiçá mesmo de introdução geral, como sugeriu Marcelo Carvalho em 2010). Isso, senão por seus acertos – que não é de ensinar acertos que vive uma professora de filosofia –, ao menos para que a formação de nossos estudantes possa ser mais completa do que quando lhes apresentamos apenas "as três grandes" escolas fundacionalistas, ignorando a voz dissonante de Wittgenstein (e, vale dizer, também a do grande Poincaré). Mostrar a nossos estudantes o exemplo deste aprendiz que buscou, de modo bastante peculiar, é verdade, contestar e ultrapassar as ideias de seus mentores pode ser um bom modo de instigá-los a fazer o mesmo e, assim, seguir estimulando o desenvolvimento da autonomia intelectual nesta longa e multifacetada conversa que é a filosofia. E talvez também a conversa entre a filosofia e a matemática.[31]

[31] A autora agradece à CAPES pelo financiamento através do AUXPE 23038.006944/2014-72 e do Projeto CAPES-Cofecub nº 88887.130207/2017-01, Processo 88887.371155/2019-00. Aos colegas Rodrigo Sabadin Ferreira, Ronai Pires da Rocha e Tamires Dal Magro, agradeço a leitura atenciosa, as correções e sugestões que melhoraram a redação deste texto.

Crédito da Imagem: Bruno Liberati.

In memoriam

(1949-2020)

Referências

AVIGAD, J. "Understanding proofs". In: MANCOSU, P. (ed) *The Philosophy of Mathematical Practice*. Oxford: Oxford University Press, 2008, pp. 317-353

CARVALHO, M. "O que se diz e o que se cala: Wittgenstein e os limites da linguagem." IN CORNELLI, G., CARVALHO, M., DANELON, M. (eds) *Filosofia: Ensino médio*. Coleção Explorando o Ensino, Vol. 14. Brasília: Ministério da Educação, Secretaria de Educação Básica, 2010, pp. 101-116.

CARTER, J. "Philosophy of Mathematical Practice – Motivations, Themes and Prospects". *Philosophia Mathematica*, Vol. 27, Issue 1, February 2019, pp.1–32.

DA SILVA, J. J. *Filosofias da matemática*. São Paulo: Editora Unesp, 2007.

D'AGOSTINI, N. *Stanislávski e o Método de Análise Ativa: a criação do diretor e do ator*. São Paulo: Perspectiva, 2018.

A Bruno Liberati, autor da caricatura acima, manifesto meu imenso carinho e a gratidão por ter generosamente cedido este seu trabalho para meu uso acadêmico em 2011. Que sua estadia em outros planos seja como você gostava de chamar as coisas boas, Bruno: *Sensacional!*

ENGELMANN, M. "Instructions for Climbing the Ladder (The Minimalism of Wittgenstein's *Tractatus*)." *Philosophical Investigations*, v. 41, pp. 446-470, 2018.

ENGELMANN, M. "What does it take to climb the ladder? (A sideways approach)". *Kriterion*, n.140, pp. 591-611, 2018.

ESQUISABEL, O. M. "¿Lenguage racional o ciencia de las formulas? La pluridimensionalidad del programa leibniziano de la Característica general" In: WRIGLEY, M. (ed.) *Manuscrito, Revista Internacional de Filosofia,* Vol. 25, nº 2: *Dialogue, Language, Rationality. A Festschrift for Marcelo Dascal*, 2002, pp. 147-197.

ESQUISABEL, O. M. "Representing and Abstracting. An Analysis of Leibniz's Concept of Symbolic Knowledge." In: LASSALLE CASANAVE, A. (ed.) *Symbolic Knowledge from Leibniz to Husserl*. London: College Publications, 2012, pp. 1-50.

FERRARI, M. & STAMATESCU, I-O. *Symbol and physical knowledge: on the conceptual structure of physics* Berlin Heidelberg New York: 2002.

FERREIRA, R. *A Lógica do Tractatus e o Operador N: Decidibilidade e Capacidade Expressiva*. Dissertação (Mestrado em Filosofia). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2017.

FERREIRÓS, J. *Mathematical Knowledge and the Interplay of Practices*. Princeton: Princeton University Press, 2016.

FLOYD, J. "Number and Ascriptions of Number in Wittgenstein's *Tractatus*" In RECK, E. (ed.), *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy* pp. New York: Oxford University Press, 2002, pp. 308–352.

FLOYD, J. "Wittgenstein on Philosophy of Logic and Mathematics. In SHAPIRO (ed.) *The Oxford Handbook of Philosophy of Mathematics and Logic*. pp. 75-128, 2007a.

FLOYD, J. "Wittgenstein and the Inexpressible". In: CRARY, A. (ed), *Wittgenstein and the Moral Life: Essays in Honor of Cora Diamond*. MIT Press, pp. 177-234, 2007b.

FRASCOLLA, P. *Wittgenstein's Philosophy of Mathematics*. London and New York: Routledge, 1994.

FRASCOLLA, P. *Understanding Wittgenstein's Tractatus*. Routledge, 2007.

FREGE, G. *Os Fundamentos da Aritmética. Uma Investigação Lógico-Matemática sobre o Conceito de Número*. Tradução de Luiz Henrique dos Santos. In. Frege/Peirce: Coleção Os Pensadores. São Paulo: Abril Cultural, 1980.

ISHIGURO, H. *Leibniz's Philosophy of Logic and Language*. London: Duckwork, 1972.

JOURDAN, C.A. *Impredicatividade, Generalidade e o Desenvolvimento do Pensamento de Wittgenstein*. Tese (Doutorado em Filosofia) – Pontifícia Universidade Católica do Rio de Janeiro. Orientador: Luiz Carlos Pereira. Rio de Janeiro, 2009.

LASSALLE CASANAVE, A. "Kant's Avatars of Symbolic Knowledge: The Role of Symbolic Manipulation in Kant's Philosophy of Mathematics." LASSALLE CASANAVE, A. (ed.), London: College Publications, 2012, pp. 51-78.

LASSALLE CASANAVE, A. "Conocimiento simbólico y aritmética en HIlbert". FERREIRÓS DOMÍNGUEZ, J. & LASSALLE CASANAVE, A. (Coord.) El árbol de los números - cognición, lógica y prática matemática. Sevilla: Editorial Universidad de Sevilla, 2016, pp. 219-234.

LASSALLE CASANAVE, A. & SAUTTER, F. T. *Visualização das Ciências Formais*. London: College Publications, 2012.

LEME, R. R. *Programação como forma de vida*: *uma crítica aao representacionalismo na teoria da computação.* Trabalho de Conclusão de Curso (Bacharelado em Filosofia). Orientador: Paulo F.E. Faria; co-orientadora: Gisele Dalva Secco. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

LEGRIS, J. "Between Calculus and Semantic Analysis". LASSALLE CASANAVE, A. (ed.), London: College Publications, 2012, pp. 79-114.

MANCOSU, P. "Introduction". In: Mancosu (ed) *The Philosophy of Mathematical Practice*. Oxford: Oxford University Press, 2008, pp. 1-21.

MARION, M. *Wittgenstein, Finitism, and The Foundations of Mathematics*. Nova York, EUA: Oxford University Press, 1998.

MARION, M. *Ludwig Wittgenstein: Introdução ao* Tractatus Logico-Philosophicus. São Paulo: Annablume Editora, 2012.

MARQUES, E. da R. *Wittgenstein e o* Tractatus. Rio De Janeiro: Zahar, 2015.

MÜHLHÖLZER, F. "'A mathematical proof must be surveyable' – what Wittgenstein mean by this and what it implies". *Grazer Philosophische Studien* vol. 71, 2005, pp. 57–86.

MUMMA, J. "The computational effectiveness of geometric diagrams". (no prelo)

NAKANO, A. L. *A matemática das Philosophische bemerkungen: Wittgenstein no contexto da* Grundlagenkrise. Tese (Doutorado). Universidade Fededral de São Carlos: UFSCar, 2015.

NARBOUX, JP. "How Showing takes care of itself". *Philosophical Topics*, v.42, n.2 2014.

POINCARÉ, J.-H. *O valor da ciência*. Tradução Maria Helena Franco Martins. Rio de Janeiro: Contraponto, 1995.

PLAUD, S. *Expression et coordination de Leibniz à Wittgenstein*. Paris: Librairie Philosophique J. Vrin, 2018.

ROCHA, R. P. da. *Ensino de filosofia e currículo*. 2ª ed. Santa Maria: Editora da UFSM, 2015.

ROCHA, R.P. da. *Escola Partida: ética e política na sala de aula*. São Paulo: Editora Contexto, 2020.

RUSSELL, B. "As funções de um professor". *Quatro textos excêntricos*. Prefácio e tradução de Olga Pombo. Lisboa: Relógio D'água, 2000.

RUSSELL, B. *Introdução à filosofia matemática*. Tradução Maria Luiz X. De A. Borges. Rio de Janeiro: Zahar, 2007

SANTOS, L.H. dos "A Essência da Proposição e a Essência do mundo." In: *Tractatus Logico- Philosophicus*. Tradução, apresentação e estudo introdutório de Luiz Henrique Lopes dos Santos, São Paulo: Editora da Universidade de São Paulo, 2001, pp. 11-128.

SANTOS, L.H. dos "Notas críticas sobre o realismo matemático, à moda de Wittgenstein". *Analytica Revista de Filosofia*, vol. 12 (1), 2008, pp.131-147.

SECCO, G.D. *Entre provas e experimentos: uma leitura wittgensteiniana das controvérias em torno da prova do Teorema das Quatro Cores*. Tese de Doutorado. Orientador: Luiz Carlos Pinheiro Dias Pereira. PUC-Rio. Rio de Janeiro, 2013.

SECCO, G.D. & NOGUEZ, P.M.R. "Operar e exibir: Aspectos do Conhecimento Simbólico na Filosofia Tractariana da Matemática". *Revista Portuguesa de Filosofia*, vol 73 (3-4), 2017, pp. 1463-1492.

SHAPIRO, S. *Filosofia da matemática*. Tradução e notas de Augusto J. Franco de Oliveira. Lisboa: Edições 70, 2015.

SILVA, M. MARION, Mathieu. *Ludwig Wittgenstein: introdução ao* Tractatus Logico-Philosophicus. São Paulo: Annablume, 2012. Princípios: Revista de Filosofia (UFRN), v. 20, n. 34, 14 jul. 2015, pp. p. 365-386.

STENLUND, S. "The "middle Wittgenstein" and Modern Mathematics" In: Dybjer et al. (eds) *Epistemology versus Ontology, Essays on the Philosophy and Foundations of Mathematics in Honour of Per Martin-Löf*. Dordrecht: Springer Science + Business Media, 2012, pp.139-159.

STENLUND, S. "Wittgenstein and Symbolic Mathematics". *O que nos faz pensar*, v.22, n.33, 2013, pp. 7-34.

STENLUND, S. *The Origin of Symbolic Mathematics and the End of the Science of Quantity*. Uppsala: Uppsala Universität, 2014.

STENLUND, S. "On the Origin of Symbolic Mathematics and Its Significance for Wittgenstein's Thought" *Nordic Wittgenstein Review*, 4, n.1, 2015, pp. 7-92.

VELLOSO, A. "Wittgenstein "Great Analysis" and Frege's construal of number as a property of properties". *Analytica*, Rio de Janeiro, vol 21 no 1, 2017, pp. 17 1-208.

WITTGENSTEIN, L. *Tractatus Logico-Philosophicus*. Tradução, apresentação e estudo introdutório de Luiz Henrique Lopes dos Santos. São Paulo: Editora da Universidade de São Paulo, 2001.

WRIGLEY, M. "Wittgenstein's Philosophy of Mathematics," *Philosophical Quarterly*, Vol. 27, n., 1977, pp. 106: 50-9.

A LÓGICA DE CLASSES DE BOOLE E A LEI INDEX[1]

Alessandro Bandeira Duarte[2]

RESUMO

No presente artigo, discutirei a afirmação de Boole segundo a qual lógica não deve ser associada à metafísica. A visão de Boole depende de uma interpretação particular da Lei Index como uma lei algébrica, da qual o princípio de contradição é derivado. Mostrarei que a visão de Boole é equivocada, uma vez que temos que rejeitar a interpretação algébrica da Lei Index. Além disso, apresentarei o sistema da lógica de classes de Boole, mostrando as vantagens e desvantagens de seu sistema. Apresentarei como derivar Barbara e Celarent a partir dos axiomas e discuto que o método de solução de equações não pode ser considerado um método de dedução.

Palavras-chave: Boole. Lei Index. lógica de Classes. Axiomas. Método de Solução de Equações.

ABSTRACT

In this paper, I will discuss Boole's claim that logic should not be associated with metaphysics. Boole's position depends on a particular interpretation of the Index Law as an algebraic law, from which the principle of contradiction is derived. I will show that Boole's view is misleading since we have to refuse the algebraic interpretation of the Index Law. In addition, I will present Boole's logic of classes, showing the advantages and disadvantages of his system. I will present how to derive Barbara and Celarent from the axioms and discuss that the method for solving equations can not be regarded a method of deduction.

Keywords: Boole. Index Law. logic of Classes. Axioms. Method for Solving Equations.

[1] Eu gostaria de agradecer à(ao) referee anônima(o) pelas valiosas contribuições ao presente artigo.

[2] Professor do departamento de filosofia da UFRRJ.
E-mail: dedekindbr@nulfic.org. ORCID: 0000-0002-4011-5000.

1. Introdução

Em 1847, Boole publicou um pequeno livro intitulado *The mathematical analysis of logic*[3], no qual ele introduziu um sistema algébrico de notação capaz de representar as proposições categóricas que participam dos silogismos. Além disso, esse sistema foi apresentado axiomaticamente, de forma que os silogismos foram derivados de leis do pensamento, que são leis algébricas[4] [5]. Em 1854, *The laws of thought* foi publicado. Nesse livro Boole melhorou e estendeu as ideias apresentadas em 1847.

Qual é o papel de Boole na história da lógica? Meu objetivo aqui é considerar a análise que Chateaubriand sustentou no artigo "Boole on Reference and Universe of Discourse: Reply to John"[6]. Segundo Chateaubriand, Boole é uma figura importante na história da lógica, não nega que Boole, de fato, pode ser considerado o iniciador da lógica moderna e que lógica algébrica de Boole desempenhou certo papel na lógica desenvolvida no fim do século 19 e início do século 20. Porém, ele considera que os trabalhos em lógica de Frege e Russell são mais essenciais, porque estes últimos consideram a lógica como uma ciência a qual trataria dos aspectos mais fundamentais da realidade.

Para Chateaubriand, o que distingue a noção de lógica de Frege e Russell da noção de lógica de Boole é a concepção metafísica[7]. Aqui temos um ponto interessante ao qual será dada uma maior atenção adiante. Boole defende que a lógica é uma ciência[8]. Ele defende que a lógica trata da relação

---

[3] O interesse de Boole pela lógica começa com a controvérsia entre William Hamilton e Augustus De Morgan em relação à autoria da teoria da quantificação de predicados. Cf. BOOLE, 1998, p. 1.

[4] Boole dividiu as proposições da lógica em dois tipos: (1) proposições primárias, que estabelecem relações entre classes e (2) proposições secundárias, que estabelecem relações entre proposições. Os axiomas algébricos estabelecidos (seção 3) por Boole valem tanto para proposições primárias como para proposições secundárias.

[5] As leis da lógica de Boole são algébricas em um sentido mais estrito. Mais adiante discutirei esse ponto.

[6] Alguns autores consideram o papel de Boole apenas secundário. Em geral, essa interpretação está atrelada à falta de uma teoria da quantificação no sistema de Boole. Veja, por exemplo, Dummett (1967, p. 66) e Heijenoort (1981, p vi). Outros autores defendem o papel proeminente de Boole na história da lógica. Dentre estes, posso mencionar Quine (1940, p. 1), Lewis e Langford (1959, p. 89), Corcoran (2003, p. 273) e Corcoran (2004, p. 163).

[7] Cf. CHATEAUBRIAND, 2004, p. 178-9.

[8] Veja, por exemplo, Boole (1998, p. 7) e Boole (2003, cap. 1).

entre classes ou da relação entre proposições. O próprio Corcoran (2003) afirma que

> Boole laid down the groundwork for science of formal laws of being, in Tarski's words 'general laws governing the concepts common to all sciences' or 'the most general laws of thinkables' (1941/1994, pp xii), to use the words that Kneale and Kneale applied to what Boole called 'laws of thought (1962/ 1988, p. 407). (CORCORAN, 2003, p. 286).

Então, por que Boole afirma que a lógica não deveria mais ser associada à metafísica? A resposta que será sugerida é a seguinte: na sua tentativa de defender que a lógica é uma espécie de álgebra, Boole foi incapaz de ver que seus princípios lógicos, em particular, a Lei Index

$x.x=x$ ou $x^2=x$

baseava-se em princípios metafísicos e não em princípios algébricos.

Na próxima seção, apresentarei e discutirei os símbolos algébricos primitivos da lógica de Boole e a proposta de tradução das proposições categóricas nessa linguagem. Na terceira seção, discutirei as leis do pensamento (axiomas) do sistema de Boole. Argumentarei que, de acordo com Boole, esses axiomas são verdadeiros quando os símbolos eletivos[9] recebem os valores 0 ou 1. Em particular, a lei Index só é algebricamente verdadeira nessa interpretação. Em minha visão, esse fato é a razão pela qual Boole rejeita o fundamento metafísico da lógica. Além disso, argumentarei que essa interpretação impede que Boole aceite algumas leis lógicas com respeito às classes, tais como

$x+x=x$ e $1+x=x$.

Finalmente, na seção 4, discutirei a derivação de algumas figuras do silogismo aristotélico dentro do sistema de Boole e abordarei alguns problemas relacionados com a classe indeterminada v e com o método de solução de equações.

[9] Na próxima seção, explico o que significa "símbolo eletivo".

2. Os símbolos primitivos e a interpretação algébrica das sentenças categóricas

2.1. Os símbolos primitivos

Não há nos dois livros de Boole já mencionados qualquer resquício do estabelecimento de uma sintaxe formal da linguagem lógica, ou seja, não encontramos nos textos dele um vocabulário inicial, nem regras sintáticas de formação de expressões bem-formadas. Entretanto, é possível reconhecer minimamente qual é o vocabulário inicial e as regras de formação. O vocabulário inicial do sistema de Boole consiste nos seguintes símbolos:

(1) o símbolo "1";
(2) o símbolo "0";
(3) os símbolos *v, v'*, etc;
(4) os símbolos "*x*", "*y*", "*z*", etc;
(5) os símbolos de operadores: "+", ".", e "–"
(6) o símbolo de relação: "=".

De acordo com Boole, símbolo "1" significa a classe universal, que contém todas as classes de indivíduos, se existentes ou não[10] [11]. Em *The mathematical analysis of logic*, Boole introduz o símbolo "0" sem muitas explicações, todavia o contexto deixa claro que se trata da classe vazia. Os símbolos "*v*", "*v'*", etc podem ter, pelo menos, duas interpretações: ora eles funcionam como classes indefinidas, ora funcionam como espécies de quantificadores que se anexam a uma determinada classe[12]. A introdução dos símbolos "*v*", "*v'*" é um dos pontos mais controversos da lógica de Boole[13]. Os

[10] Cf. BOOLE, 1998, p. 15.

[11] Na verdade, Boole muda sua interpretação de "1" como sendo o universo de discurso fixo (em *Mathematical Analysis of Logic*) para um universo de discurso variável (em *The Laws of Thought*). Mas para facilitar esta exposição desconsiderarei este fato.

[12] Boole (1998, p. 39) introduz explicitamente *v* e *v'* para traduzir as sentenças "Alguns Xs são Ys" e "Alguns Ys são Zs". O mesmo procedimento é adotado em Boole (2003, p. 231).

[13] Talvez um dos objetivos de Boole ao introduzir a classe indeterminada "*v*", como no caso da introdução da classe nula, seria uniformizar a interpretação lógica de sentenças contendo a expressão "Algum X". O que queremos dizer é o seguinte: Boole poderia perfeitamente traduzir "Algum X é Y" por "$x.y \neq 0$". Contudo, Boole deseja interpretar todas as sentenças como identidades com o objetivo de encontrar um método de dedução.

símbolos "*x*", "*y*", "*z*", etc são símbolos chamados *eletivos*, uma vez que eles "elegem" do universo (classe universal) um certo subconjunto. Assim, "*x*" refere-se à classe dos indivíduos do universo que tem a propriedade X, "*y*", à classe dos indivíduos do universo que tem a propriedade Y e assim por diante.

"+", "." e "–" são operadores que agem sobre as classes, produzindo novas classes. "+" funciona como uma espécie de união entre classes; ".", como uma espécie de interseção; e "–", como uma espécie de complemento relativo. Entretanto, "+" não representa exatamente a operação de união, pois ela só produzirá um termo, se as classes forem disjuntas. Em relação ao símbolo "–", há também uma restrição, a saber, a classe no subtraendo tem de ser uma subclasse da classe no minuendo. Finalmente, "=" ("é o mesmo que") representa uma relação entre classes (nos silogismos, trata-se da predicação).

Dados os primitivos lógicos acima, podemos agora afirmar regras sintáticas para produzir expressões bem-formadas na linguagem de Boole. As seguintes regras indutivas sugerem-se:

I. "1", "0", "*v*", "*v'*", etc, "*x*", "*y*", "*z*", etc são termos (caso-base);
II. se A e B[14] são termos, então A.B é um termo[15];
III. Se A e B são termos e A.B=**0**[16], então A + B é um termo;
IV. Se A e B são termos e B⊆A, então A – B é um termo;
V. Todo termo é formado a partir das regras I-IV;
VI. Se A e B são termos, então A=B é uma fórmula (equação);
VII. Toda fórmula é formada pela regra VI.

Note que não apresentei nenhuma expressão que denotasse a negação. De fato, Boole não tinha tal expressão. Não obstante, na linguagem de Boole, é completamente possível formar termos negativos. Para isso, Boole utiliza o seguinte artifício: uma vez que "1" denota a classe de todas as coisas, e assu-

[14] Aqui "A" e "B" são metavariáveis que percorrem os termos da linguagem-objeto.

[15] Os operadores representam a si mesmo na metalinguagem. Faço isso para evitar a introdução de aspas ou novos símbolos.

[16] "**0**" refere-se à classe vazia na metalinguagem. O que pretendo expressar é que as classes devem ser disjuntas para que a união entre elas represente também uma classe.

mindo que "x" denota a classe dos Xs, então "$1-x$" denotaria a classe das coisas que não são Xs, em outras palavras, "$1-x$" pode ser interpretado como o termo negativo "não-X"[17]. Além disso, é possível mostrar, dada esta interpretação de Boole do termo negativo "não-X", que o par de termos contraditórios exaure o universo. Seja "x" a classe dos Xs, seja "$1-x$" a classe dos não-Xs, então $x+(1-x)$[18] é igual a 1, ou seja, o universo.

2.2 A interpretação algébrica das sentenças categóricas

Agora, será analisada a interpretação lógica de Boole das sentenças usadas nos silogismos. Aristóteles, em *Prior Analytics*, trabalha com quatro tipos de sentenças, a saber:

A. a sentença universal afirmativa (Todo X é Y): A(x,y);
E. a sentença universal negativa (Nenhum X é Y): E(x,y);
I. a sentença particular afirmativa (Algum X é Y): I(x,y);
O. e a sentença particular negativa (Algum X não é Y): O(x,y)[19].

Para Boole, só há um tipo de conector "é mesmo que", representado pelo símbolo "=". Este fato ocorre porque Boole interpreta "Todo X", "Nenhum X"[20], "Algum X" como classes, de forma que "Todo X é Y" significa "a classe inteira X é Y"[21]. O "é" representa a identidade entre a classe inteira X e uma subclasse da classe Y.

Assim, na linguagem simbólica de Boole, as quatro sentenças acima são interpretadas da seguinte forma:

---

[17] Certamente a classe X é um subconjunto da classe universal, então "$1-x$" é um termo.

[18] A classe dos Xs e a classe dos não-Xs são disjuntas, assim "$x+(1-x)$" é um termo.

[19] Essas abreviações serão usadas mais adiante.

[20] Na verdade, "Nenhum X" não denota uma classe propriamente dita. Boole transforma a sentença na qual ocorre esta expressão (universal negativa) em uma sentença onde tal expressão desaparece. Por exemplo, "Nenhum X é Y" significa a mesma coisa que "Todo X é não-Y" (por obversão). Esta última sentença é aceitável para Boole, porque "Todo X" denota a classe dos Xs e "não-Y" pode ser perfeitamente expresso na sua linguagem. Cf. BOOLE, 2003, p. 228.

[21] Cf. BOOLE, 2003, p. 228.

A'. Todo X é Y: $x.(1-y) = 0$[22] ou $x=x.y$[23][24]

E'. Nenhum X é Y: $x.y = 0$[25][26]

I'. Algum X é Y: $x.y = v$[27]

O'. Algum X não é Y: $x.(1-y) = v$[28]

Aqui, temos de fazer algumas observações. A princípio, parece que Boole não assume a quantificação de predicados, mas isto está implicitamente assumido nas fórmulas. Por exemplo, "$x=x.y$" implica que x é uma subclasse de y e assim a classe dos x é igual a alguma subclasse da classe dos y. Portanto, implicitamente, "*Y*" é entendido em (A') como "Algum *Y*"[29]. Caso contrário, (A') afirmaria a extensionalidade entre as classes dos Xs e Ys ($x = y$). De fato, se ocorrer que $x=x.y$ e $y=y.x$, então $x=y$. A notação acima foi apresentada por Boole (1998). Já em *The Laws of Thought*, Boole explicitamente apresenta a quantificação de predicados. (A') é expressa lá da seguinte forma:

A". Todo X é Y: $x=v.y$

Na verdade, Boole (2003, p. 228) assume explicitamente, na sua notação, a quantificação do predicado para as sentenças de (A')–(O'). Ele escreve:

> The expressions of the eight fundamental types of proposition in the language of symbols are as follows:
>
> | | |
> |---|---|
> | 1. All Y´s are X´s | $y=vx$ |
> | 2. No Y´s are X´s | $y=v(1-x)$ |
> | 3. Some Y´s are X´s | $vy=vx$ |
> | 4. Some Y´s are not-X´s | $vy=v(1-x)$ |
> | 5. All not-Y´s are X´s | $1-y=vx$ |
> | 6. No not-Y´s are X´s | $1-y=v(1-x)$ |

[22] A interseção entre a classe dos *X*s e a classe dos não-*Y*s é (igual a) a classe nula.

[23] A classe dos *X*s é (igual a) a interseção da classe dos *X*s e da classe dos *Y*s.

[24] Cf. BOOLE, 1998, p. 21.

[25] A interseção da classe dos *X*s e da classe dos *Y*s é (igual a) a classe nula.

[26] Cf. BOOLE, 1998, p. 21.

[27] A interseção da classe dos *X*s e da classe dos *Y*s é (igual a) uma classe indeterminada de *V*s que pertencem às classes dos *X*s e dos *Y*s. Cf. BOOLE, 1998, p. 21.

[28] A interseção da classe dos Xs e dos não-Ys é (igual a) uma classe indeterminada de Vs que pertencem às classes dos Xs e dos não-Ys. Cf. BOOLE, 1998, p. 22.

[29] Há passagens que sugerem esta interpretação em *The Mathematical Analysis of Logi*c. Cf. BOOLE, 1998, p. 21. Veja também Boole (1998, p. 24).

7. Some not-Y´s are X´s $v(1-y)=vx$
8. Some not-Y´s are not-X´s $v(1-y)=v(1-x)$.

Esta passagem sugere que o termo-predicado "X" tem de ser entendido por "Algum X"[30].

Boole (1998) estende a concepção das sentenças que podem ser usadas nos silogismos admitindo termos-sujeito negativos. Além disso, ele admite termos complexos (tanto termo-sujeito como termo-predicado). Assim é perfeitamente possível expressar a sentença "Todas as baleias que são azuis são mamíferas" na notação de Boole:

$x.y(1-z)=0$ ou $x.y= v.z$,

onde "x" significa a classe das baleias, "y", a classe das coisas azuis e "z", a classe dos mamíferos[31].

3. As leis do pensamento

Na introdução, afirmei que Boole defende a ideia de que a lógica é uma ciência. Neste sentido, ele poderia ser colocado ao lado de Frege, Russell e Gödel. É interessante que um dos pontos do artigo de Chateaubriand (2004) – em resposta ao artigo de Corcoran (2004) – é que ele defende uma concepção metafísica de lógica (lógica como ciência), diferentemente de Boole (e talvez de Corcoran). Aqui, acredito que Chateaubriand foi enganado pelas palavras de Boole quando ele diz que "we ought no longer to associate Logic and Metaphysic" (BOOLE, 1998, p. 13). De fato, Chateaubriand deve ser isentado do engano, porque o próprio Boole não pôde perceber que seus princípios lógicos não eram matemáticos, mas metafísicos. Mas antes de entrar neste assunto, discutirei um pouco mais a visão de Boole de lógica como ciência.

Boole (1998, p. 7) escreve o seguinte:

[30] Em seus escritos póstumos a mesma visão é defendida. Cf. BOOLE, 1997, p. 15.
[31] Cf. BOOLE, 2003, p. 35.

> It would be premature to speak of the value which this method may possess as an instrument of scientific investigation. I speak here with reference to the theory of reasoning, and to the principle of a true classification of the forms and cases of Logic considered as a Science.

Mas, para Boole, quais são os objetos de estudo da lógica? Chateaubriand defende, em seu livro *Logical Forms*, que a lógica é uma teoria ontológica de propriedades, em particular, das propriedades lógicas. Para Frege, lógica é uma teoria ontológica de objetos (lógicos) e propriedades. Para Boole, os objetos de estudo da lógica são as classes e suas relações e as proposições e suas relações. Em *The Mathematical Analysis of Logic*, Boole não expressou explicitamente isso, mas, em *The Laws of Thought*, ele escreve:

> Logic is conversant with two kinds of relations, - relations among things, and relations among fact. But facts are expressed by propositions, the latter species of relation may, at least for the purposes of Logic, be resolved into a relation among propositions. The assertion that the fact or event A is an invariable consequent of the fact or event B may, to this extent at least, be regarded as equivalent to the assertion, that the truth of the proposition affirming the occurrence of the event B always implies the truth of proposition affirming the occurrence of the event A. Instead, then, of saying that Logic is conversant with relations among things and relations among facts, we are permitted to say that it is concerned with relations among things and relations among propositions (BOOLE, 2003, p. 7).

É claro, toda ciência tem suas leis. E dentre estas leis, há aquelas que são mais fundamentais (os axiomas), a partir das quais todas as demais são derivadas. Boole está consciente deste fato[32].

Boole sustenta que é uma função lógica o ato de conceber classes. Além disso, como ciência, a lógica tem de ter suas próprias leis, dentre as quais há as leis mais básicas. Boole propõe o seguinte conjunto de axiomas lógicos (as leis do pensamento):

I. x=x

II. x.y=y.x

III. $x^2=x$ ($x^n=x$)[33]

[32] Cf. BOOLE, 2003, p. 5.

[33] De acordo com Boole, a Lei Index é uma lei lógica relacionada às classes, uma vez que o ato de selecionar os Xs da classe universal e novamente selecionar os Xs dessa classe resultante produz novamente a classe dos Xs.

IV. x+y=y+x

V. $z.(x+y)=z.x+z.y$

VI. $z.(x-y)=z.x-z.y$

VII. $1.x=x$

VIII. $0.x=0$

Estas leis são explicita ou implicitamente afirmadas em, pelo menos, um dos dois livros de Boole. (1) é a lei da identidade; (2) afirma que a comutatividade da operação de multiplicação (interseção) de classes; (3) é a *Lei Index*; (4) afirma a comutatividade da operação de adição (união) de classes; (5) é a lei distributiva da multiplicação em relação à soma (união) de classes; (6) é a lei distributiva da multiplicação em relação à subtração de classes; (7) afirma que a interseção entre a classe universal e qualquer outra classe resulta nesta mesma classe; e (8) afirma que a interseção entre a classe nula e qualquer outra classe resulta na classe nula.

Boole também tem algumas regras de inferências as quais chamarei, seguindo Corcoran (1980):

Regra da funcionalidade da Adição (RFA): $x=y, z=w \vdash x+z=y+w$
Regra da Funcionalidade da Subtração (RFS): $x=y, z=w \vdash x-z=y-w$
Regra da Funcionalidade da Multiplicação (RFM): $x=y, z=w \vdash x.z=y.w$

Casos particulares das regras acima são, respectivamente:

(RFA'): $x=y \vdash x+z=y+z$
(RFS'): $x=y \vdash x-z=y-z$
(RFM'): $x=y \vdash x.z=y.z$[34]

Segundo Corcoran (1980, p. 616), Boole implicitamente sugeriu como regras a simetria e a transitividade da identidade.

[34] Cf. BOOLE, 2003, p. 36.

(SI): $x=y \vdash y=x$

(TI): $x=y,\ y=z \vdash x=z$

E também, em uma nota na página 18 de *The mathematical analysis of logic*, Boole menciona as regras:

(IgT): $x=z,\ y=z \vdash x=y$

(DiT): $x=z,\ y\neq z \vdash x\neq y$

Contudo, (**DiT**) não tem qualquer utilidade para Boole, uma vez que desigualdades não são expressas em seu sistema[35].

Há no conjunto de axiomas de Boole algumas omissões de leis que poderiam ser consideradas axiomas lógicos. Em primeiro lugar, Boole não assume explicitamente nem a associatividade da multiplicação, nem a associatividade da adição de classes. De acordo com Corcoran (1980, p. 616), "...perhaps it was his casual attitude concerning parentheses which led him [Boole] to overlook the two associative laws".

Contudo, há outras leis que não foram afirmadas por Boole, porque, acredito, elas iriam de encontro com a sua ideia de que a lógica é uma espécie de matemática (uma álgebra da mente). Em uma passagem citada acima, Boole afirma que a lógica tem de ser associada à matemática e não mais à metafísica. Essa afirmação é sustentada pelo seguinte fato, a saber, da Lei Index, Boole deriva uma fórmula que pode ser interpretada como o *Princípio de Contradição*. Ele escreve:

> *That axiom of metaphysicians which is termed the principle of contradiction, and which affirms that it is impossible for any being to possess a quality, and at the same time not possess it, is a consequence of the fundamental law of thought, whose expression is* $x^2=x$.
>
> Let us write this equation in the form
>
> $$x-x^2=0$$

[35] Boole não tem símbolo para negação de fórmulas. Além disso, "≠" não é primitivo no sistema.

> whence we have
>
> $$x.(1-x)=0;$$
>
> both these transformations being justified by the axiomatic laws of combination and transposition (II.13) (BOOLE, 2003, p. 49) [36].

Ora, em que sentido Boole acredita que a lógica não depende mais da metafísica? Para Boole, os axiomas lógicos apresentados acima são axiomas algébricos[37]. Em um artigo póstumo, Boole escreve o seguinte:

> [1st] It has been shewn that if x and y are any two elective symbols then
>
> $$x+y=y+x.$$
>
> Now the same equation holds true if x and y represent numbers. We know 7+5=5+7, 6+2=2+6 and generally whatever numerical values we attach to x and y that $x+y=y+x$.
>
> 2nd. It has also been shewn that x and y being any elective symbols we have
>
> $$x.y=y.x$$
>
> and this relation is equally true if x and y are any arithmetical symbols. Thus 7×5=5×7, 8×9=9×8 and similarly for any other values of x and y…
>
> 4th. Any elective symbol x satisfies the law whose expression is
>
> $$x^2=x.$$
>
> It is not true that any numerical symbol satisfies this law. Out of the entire range of numerical magnitude from minus infinite to plus infinity there are but two magnitudes which satisfies this law they are 0 and 1. For we have
>
> $$0\times0=0 \text{ or } 0^2=0$$
>
> and
>
> $$1\times1=1 \text{ or } 1^2=1.$$
>
> These numerical magnitudes may in fact be introduced into the elective system and admit of interpretation there as well as in the system of number to which they also belong (BOOLE, 1997, p. 7-8).

[36] Embora Boole não tenha mencionado este fato em *The Mathematical Analysis of Logic*, pode ser que ele já tivesse isso em mente.

[37] Cf. BOOLE, 2003, p. 50.

De acordo com essa passagem, podemos dividir os axiomas em duas classes, a saber, aqueles que são válidos para quaisquer valores numéricos e aquele que é válido para valores numéricos restritos. Formam a primeira classe os axiomas 1), 2), 4), 5), 6), 7) e 8). A segunda classe é formada justamente pelo axioma 3), a Lei Index, cuja validade se dá para valores 0 ou 1: $0^2=0$ ($0^n=0$) e $1^2=1$ ($1^n=1$).

Em *Laws of Thought*, Boole também defende que a sua álgebra de classe é uma álgebra que pode ser interpretada, tomando-se os valores 0 e 1 para os símbolos eletivos[38]. É nesse sentido que sua *Lei Index* é uma lei algébrica e, portanto, que a Lei de Contradição é uma lei algébrica (derivada) e não um axioma metafísico.

Agora, podemos entender porque no sistema de lógica de Boole, não há os axiomas

1) $x+x=x$ ou $2.x=x$;
2) $1+x=x$[39].

Suponha que "x" tome os valores 0 ou 1. As leis acima serão verdadeiras somente se "x" for igual a 0[40]. Portanto, parece-nos, é totalmente justificado, dada a interpretação de Boole, a não-inclusão destas leis. Para mim, elas não foram simples omissões. Aceitá-las significaria dizer que não havia nenhuma analogia entre os axiomas lógicos e as leis algébricas e, portanto, a redução da lógica à matemática estaria fadada ao fracasso.

É interessante mencionar que, para Boole, a Lei Index não funciona quando aplicada a uma adição de classes. Por exemplo, a equação

$(x+y)^2=x+y$

[38] Cf. BOOLE, 2003, p. 37-8, 70, 72.

[39] Corcoran (1980, p. 617) sustenta que a ausência de $1+x=1$ e $x+x=x$ são meras omissão e que se poderia adicionar essas leis. Como tentei argumentar, isso não é possível para Boole. Corcoran também afirma que não há em *The mathematical analysis of logic* a lei de não-contradição. Ora, essa lei é uma consequência direta da Lei Index. Portanto, para mim, não se trata exatamente de uma omissão. Corcoran ainda menciona a omissão das leis de De Morgan. Aqui, acredito, essas leis também não poderiam ser válidas na interpretação de Boole, pois "$x+y$" só produzirá um termo, se as classes forem disjuntas.

[40] Na verdade, sintaticamente "$x+x$" não pode ser termo. Logo, "$x+x=x$" não é fórmula.

não é válida no sistema de Boole. Pois, fatorando, $(x+y)^2$ é igual a

$x^2+2.x.y+y^2$.

Como $x^2=x$ e $y^2=y$, temos

$x+2.x.y+y$.

Todavia, "$2.x.y$" não pode ser eliminado na interpretação lógica no sistema de Boole.

Porém, se assumirmos (1) e (2), a igualdade vale. Por (1), temos que

$x+x.y+y$.

Ora, daí é obtido que

$x.(1+y)+y$.

E por meio da fórmula (2), temos

$x.1+y$.

Mas pelo axioma (7), temos então

$x+y$.

A Lei Index vale para termos complexos produzidos por meio da multiplicação de classes. Assim, em geral, o seguinte vale:

3) $(x.y)^2=x.y$

De $(x.y)^2$, obtemos $x^2.y^2$. Mas $x^2=x$ e $y^2=y$. Logo, $x.y$.

A Lei Index também vale para o seguinte termo: $1-x$.

4) $(1-x)^2=1-x$

De $(1-x)^2$, obtemos $1^2-2.x+x^2$. Mas $1^2=1$ e $x^2=x$. Assim, temos $1-2.x+x$. Ou seja, $1-x$[41].

De acordo com a minha interpretação dada nessa seção, Boole confundiu-se. A *Lei Index* e o Princípio de Contradição são equivalentes. Na sua ânsia de eliminar a interpretação metafísica do último, Boole direcionou sua álgebra da lógica para um caminho tortuoso. Por isso, ele não pôde expressar leis lógicas que são bastante naturais. De fato, os sucessores de Boole logo assumiram (1) e (2).

4. O tratamento de Boole dos silogismos

Um dos objetivos de Boole em *The Mathematical Analysis of Logic* (e também em *The laws of thought*) é deduzir, dentro da sua linguagem lógica, os silogismos até então aceitos, bem como mostrar a invalidade de certos argumentos. Vale mencionar o que é um silogismo. De acordo com Aristóteles, um silogismo é formado por um conjunto de proposições a partir do qual uma outra proposição (a conclusão) necessariamente se segue.

Aristóteles não impõe nenhuma restrição no número de proposições utilizadas em um silogismo. Não obstante, em *Prior Analytics*, Aristóteles menciona apenas argumentos compostos de três proposições, sendo duas premissas e uma conclusão. As deduções de Aristóteles dos silogismos são executadas usando-se oito regras, as quais três são regras de conversão, quatro são silogismos perfeitos e uma é a regra de redução ao absurdo[42]. As regras de conversão são: **regras de conversão total (ou simples), regra de conversão parcial.**

[41] Também vale para o termo "x+(1−x)": $[x+(1-x)]^2=x+(1-x)$.

[42] A prova por redução ao absurdo é necessária, quando uma das premissas é particular negativa. Aristóteles também usa *ecthesis*, mas não entraremos em detalhes.

Regra de Conversão Simples A (RCT A): $E(x,y) \vdash E(y,x)$

Regra de Conversão Simples B (RCT B): $I(x,y) \vdash I(y,x)$

Regra de Conversão Parcial (RCP): $A(x,y) \vdash I(y,x)$

Os silogismos perfeitos são os silogismos da primeira figura[43] que são:

Barbara: $A(x,y), A(y,z) \vdash A(x,z)$

Celarent: $A(x,y), E(y,z) \vdash E(x,z)$

Darii: $I(x,y), A(y,z) \vdash I(x,z)$

Ferio: $I(x,y), E(y,z) \vdash O(x,z)$

Agora, há elementos suficientes para deduzir outros silogismos. Por exemplo, é possível deduzir o seguinte silogismo da segunda figura:

Cesare: $E(x,y), A(z,y) \vdash E(z,x)$

Para provar isso, basta aplicar (RCT A) em $E(x,y)$. Então é obtido $E(y,x)$. De $E(y,x)$ e $A(z,y)$, de acordo com *Celarent*, é obtido $E(z,x)$.

É importante também mencionar a maneira pela qual Aristóteles exclui certos argumentos como silogísticos. Aristóteles afirma, como regra geral para a segunda figura do silogismo, que uma das premissas tem de ser negativa. Se ambas as premissas forem afirmativas, nenhuma conclusão seguir-se-á por necessidade. Por exemplo, assuma o seguinte argumento como silogístico:

$A(x,y), A(z,y) \vdash A(z,x)$

É possível mostrar que as premissas podem ser verdadeiras e que a conclusão pode ser falsa. Suponha que "*x*", "*y*" e "*z*" sejam, respectivamente, "homem", mamífero" e "baleia". Neste caso, as premissas são verdadeiras,

[43] A divisão das figuras do silogismo em primeira, segunda, terceira e quarta tem a ver com a posição do termo médio, o termo que é eliminado da conclusão. Na primeira figura, o termo médio ora termo-sujeito, ora termo-predicado. Na segunda figura, o termo médio aparece como termo-predicado nas duas premissas; e na terceira figura, o termo médio aparece como termo-sujeito nas duas premissas. A quarta figura é parecida com a primeira, mas os termos na conclusão são trocados.

mas a conclusão é falsa. É desta forma que Aristóteles mostra a invalidade de argumentos[44].

Dada esta pequena exposição do sistema de dedução de Aristóteles, voltemo-nos a Boole. Um dos pontos de distinção do sistema de dedução de Boole em relação ao de Aristóteles é o uso de axiomas lógicos. De acordo com Corcoran (1980, p. 615), "this is a striking qualitative improvement on Aristotle's direct deductions which were constructed using eight rules...but no logical axioms". Agora, entenderemos por que Boole expressa todas as suas sentenças na forma de uma identidade. Como observado acima, ocorre, nos silogismos, um termo (chamado termo médio) que desaparece na conclusão. Boole tem de ter alguma forma de excluir tal termo no seu sistema de dedução. Ora, algebricamente, há uma maneira de se fazer isto. Por exemplo, sejam as equações

1. $4x+3y+2z = 0$
2. $x+2y+3z = 0$

Se quisermos excluir x do sistema, basta multiplicarmos a segunda equação por -4. Assim, obtemos

2'. $-4x-8y-12z=0$.

E somamos a primeira equação por esta última, resultando em

3. $-5y-10z = 0$ ou $5y+10z = 0$.

Vale mencionar que, nesta derivação, outros passos implicitamente foram dados. Em particular, é assumido que

$-4 = -4$ (identidade).

[44] Na verdade, o método de Aristóteles é mais completo, refutando qualquer proposição categórica possível como consequência lógica dessas premissas. Mas não entrarei em detalhes.

Também é assumida a regra RFM', tomando-se para "x" a expressão "x+2y+3z" e para "y", "0".

Boole apresenta o seguinte método geral de eliminação de um termo (análogo ao exemplo acima):

Eliminação 1 (E1)[45]:

$$a.y+b=0$$
$$a'.y+b=0$$
$$a.b'-a'.b=0$$

Entretanto, dada a solução de equações, esse esquema de eliminação de termos pode ser substituído, por uma questão de conveniência e praticidade, pelo seguinte outro esquema:

Eliminação 2 (E2)[46]

$$b=a.y$$
$$a'.y=b'$$
$$a'.b=a.b'$$

De acordo com Corcoran (1980), tanto **E1** como **E2** são válidas no sistema lógico de Boole. Seguindo-o, apresentarei a dedução de **E2**[47]:

1. $b=a.y$	premissa
2. $a'.y=b'$	premissa
3. $a'=a'$	Identidade
4. $a'.b=a'.(a.y)$	1,3 **RFM**
5. $a=a$	Identidade
6. $a.(a'.y)=a.b'$	2,5 **RFM**
7. $a'.(a.y)=(a'.a).y$	Associatividade da Multiplicação
8. $a.(a'.y)=(a.a').y$	Associatividade da Multiplicação
9.$(a'.a).y=a'.(a.y)$	7, simetria da identidade

[45] Veja Boole (1998, p. 32).

[46] Veja Boole (1998, p. 34-5).

[47] A derivação será um pouco mais completa em relação à de Corcoran.

10. $a'.b=(a'.a).y$	9,4, **IgT**
11. $(a.a').y=ab'$	8,6 **IgT**
12. $a.a'=a'.a$	Comutatividade da Multiplicação
13. $y=y$	Identidade
14. $(a.a').y = (a'.a).y$	12,13, **RFM**
15. $a'.b=(a.a').y$	14,10, **IgT**
16. $a.b' = (a.a').y$	11, Simetria da Identidade
17. $a.b'=a'.b$	16,15, **IgT** (QED)

Dado este sistema de eliminação (**E2**), Boole deduz algumas figuras dos silogismos. Por exemplo, o modo **Barbara** é derivado da seguinte forma:

Todo X é Y A(x,y)	$x.(1- y) = 0$ ou $x = x.y$
Todo Y é Z A(y,z)	$y.(1- z) = 0$ ou $y = y.z$

Em primeiro lugar, podemos resolver ambas as equações, obtendo-se: $x=v.y$ e $y=vz$[48] [49]. Agora, podemos aplicar **E2**, ou seja, multiplicando as equações "$x = v.y$" e "$y = v.z$" e eliminando "y", obtemos

$$x = v.z,$$

que expressa a sentença "Todo X é Z" (A(x,z)).

É interessante mencionar que Boole assume os axiomas lógicos, mas na dedução dos silogismos, ele só usa **E2** (ou **E1**). Porém, como pode ser visto na dedução de **E2**, muitos axiomas lógicos são usados. De fato, é possível dar uma dedução completa de *Barbara* usando, em vez de **E2**, os axiomas lógicos:

1. $x=x.y$	Premissa
2. $y=y.z$	Premissa
3. $z=z$	Axioma 1
4. $z.x=z.(x.y)$	1, 3 **RMF**
5. $z.(x.y)=(z.y).x$	4, Associatividade

[48] Como já mencionado, essa já seria a interpretação padrão em Boole (2003).

[49] Mais adiante, discuto a solução de equações.

6. $z.y=(z.y).x$	4,5, **IgT**
7. $x=x$	Axioma 1
8. $y.x=(y.z).x$	7,8 **RMF**
9. $y.x=(z.y).x$	8, Comutatividade
10. $z.x=y.x$	6, 9, **IgT**
11. $z.x=x.y$	10, Comutatividade
12. $zx=x$	1, 11, **IgT**
13. $x=x.z$	12, Comutatividade

O modo **Celarent** é derivado da seguinte forma, usando-se E2:

1. Todo X é Y A(x,y)	$x = x.y$	Premissa
2. Nenhum Y é Z E(y,z)	$y.z=0$	Premissa
3. Nenhum X é Z E(x,z)	$x.z=0$	1, 2 **E2**

A derivação completa de **Celarent** sem usar E2 pode ser executada da seguinte forma:

1. $x=x.y$	Premissa
2. $y.z=0$	Premissa
3. $x=x$	Axioma 1
4. $x.(y.z)=0$	2,3, **RMF**
5. $z=z$	Axioma 1
6. $z.x=z.(x.y)$	1, 5, **RMF**
7. $x.z=x.(y.z)$	6, Comutatividade e Associatividade
8. $x.z=0$	4, 7, Transitividade

De acordo com Corcoran (1980), há três problemas no sistema de dedução de Boole. Em primeiro lugar, diferentemente de Aristóteles, Boole não tem nenhum tipo de dedução indireta (via redução ao absurdo). Segundo Corcoran, uma possível justificativa para isso é a falta de uma expressão que funciona como negação (no seu sentido próprio). Esse é um ponto interessante, mas não entrarei nos detalhes. O segundo problema é o que Corcoran (1980, 2003) chama de "falácia das soluções" (*fallacy of solutions*). Esta falácia está intimamente ligada ao símbolo "*v*" que representa a classe indeterminada.

Boole mostra que as conversões (de sentenças) apresentadas por Aristóteles são deriváveis em seu sistema. A conversão simples é obtida, usando-se o axioma da comutatividade da multiplicação. Por exemplo[50]:

De “Nenhum X é Y”

$x.y=0$,

obtemos

$y.x=0$;

de “Algum X é Y”

x.y=v

obtemos

y.x=v,

isto é, “Algum Y é X”.

O problema encontra-se na conversão parcial. Boole escreve:

> The equations A and E, written in the forms
>
> $$(1-y) = 0$$
> $$y.x = 0$$
>
> give on solution the respective forms
>
> $$x = vy$$
> $$x=v(1-y)$$
>
> the correctness of which may be shewn by substituting these values of x in the equations to which they belong, and observing that those equations are satisfied quite independently of the nature of the symbol *v*” (BOOLE, 1998, p. 27).

Para entendermos esta passagem, é necessário entrar na questão “das propriedades das funções eletivas”[51] (BOOLE, 1998, p. 60-81) ou “dos princípios do raciocínio simbólico” (BOOLE, 2003, p. 66-79).

Partindo do teorema Maclaurin[52], Boole chega ao seguinte resultado: qualquer que seja a função de um símbolo eletivo, temos que

[50] Veja Boole (1998, p. 26).

[51] Uma função eletiva é uma função em termos dos símbolos eletivos: *x, y, z,* etc.

[52] Este teorema é: $\phi(x)=\phi(0)+\phi'(0)\,x+\frac{\phi''(0)}{1.2}\,x^2+\ldots$.

$f(x)=f(1)x+f(0)(1-x)$.

Este resultado é mais facilmente obtido em *The laws of thought*. Assuma uma função qualquer de um símbolo eletivo

$f(x)=ax+b(1-x)$[53].

Se x for 0, temos

$f(0)=b$.

Se x for 1, temos

$f(1)=a$.

Substituindo na função original "a" por "$f(1)$" e "b" por "$f(0)$", obtemos

$f(x)=f(1)x+f(0)(1-x)$ (teorema **T1**).

Mas qual é importância deste teorema? Ora, podemos considerar a fórmula

$x.(1-y) = 0$ (Todo X é Y)

da seguinte forma:

$$x=\frac{0}{1-y}.$$

[53] Esta expressão é obtida via definição. Cf. BOOLE, 2003, p. 72.

E podemos expandir, usando **T1**, a expressão "$\frac{0}{1-y}$" (considerada como uma função de *y*). Assim, para *y*=1, temos

$$\frac{0}{0}\,y;$$

para *y*=0, temos

$$\frac{0}{1}(1-y).$$

Na interpretação de Boole, "$\frac{0}{1}$" representa a classe vazia. E pelo axioma 8, $0.(1-y)=0$. Seguindo a interpretação aritmética que afirma que a divisão de zero por zero é indeterminada, Boole considera que "$\frac{0}{0}$" representa uma classe indeterminada – a classe *v*. Assim, de

$$x.(1-y) = 0,$$

Boole deriva a equação

$$x=\frac{0}{0}\,y \text{ ou } x = v.y.$$

De acordo com Corcoran (1980), o erro de Boole reside no fato dele acreditar que um método de obtenção de solução para uma determinada função é o mesmo que uma dedução.

Vale salientar um problema que é observado e que tem relação com isso. Como disse anteriormente, Boole assume em *The laws of thought* que a fórmula que expressa a sentença "Todo X é Y" é

$$x = v.y.$$

Nesta fórmula, Boole explicitamente afirma a quantificação do predicado. Boole (2003, p. 229) mostra que a fórmula acima implica a fórmula que ele usou em *The mathematical analysis of logic*, a saber,

$x = x.y$ ou $x.(1 - y) = 0$.

Para isto, basta multiplicar "$x = v.y$" por "$1 - y$":

$x(1 - y) = v.y(1 - y)$[54]

Ora, à direita da identidade, obtemos

$v.y - vy^2$.

Mas, pela Lei Index, temos

$v.y - vy$[55].

Deste modo, somos levado à seguinte conclusão: se de $x.(1 - y)=0$, obtemos $x=v.y$ (assumindo que isto ocorre por meio de uma dedução lógica); e se de $x=v.y$, obtemos $x.(1-y)=0$ (por dedução lógica), então ambas as fórmulas deveriam ser equivalentes. Isso significa o seguinte: todas as conclusões que são derivadas de uma são derivadas da outra e vice-versa. Entretanto, esta afirmação é incorreta! Por exemplo, Boole (1998, p. 35) deriva a conclusão do silogismo da quarta figura (*Fesapo*) usando a fórmula

$y=vx$.

[54] Aqui, alguns passos onde são utilizados axiomas lógicos e regras de inferência foram omitidos.

[55] Do axioma 1, temos: $x=x$. Assim, podemos obter o seguinte: $x-x=0$.

A derivação ocorre da seguinte forma:

Nenhum Z é Y	$0=zy$
Todo Y é X	$y=vx$

Usando **E2**, obtemos

$0=zvx$[56],

que significa que *alguns Xs não são Zs*. Aqui vale mencionar a segunda interpretação do símbolo "*v*". "*v*" ocorre, nesta dedução, anexado ao símbolo "*x*", de maneira que não poderíamos entender por esta fórmula que "Alguns Zs não são Xs". "*v*" parece funcionar como uma espécie de quantificador[57].

Contudo, se, em vez de $y=vx$, usarmos a fórmula

$y(1-x)=0$,

obtemos, usando **E2**, o seguinte

$0=0$[58] [59].

Esta equação, como será observado abaixo, é usada por Boole para mostrar a invalidade de um argumento. Portanto, estas duas fórmulas não podem ser equivalentes. Isto significa que o método de conversão parcial de Boole não é uma dedução lógica.

Antes de passarmos para o terceiro e último problema, gostaria de fazer algumas especulações. Se Boole assumisse que a sentença "Todo X é Y" é expressa apenas pela fórmula "$x=v.y$" e rejeitasse seu método de conversão parcial, o sistema seria interessante? Por um lado, penso que sim! Como foi visto, a partir desta fórmula, podemos derivar a fórmula "$x(1-y)=0$" ou

[56] Para aplicar E2, considere: $b=0$; $a=z$; $a'=1$; e $b'=v.x$.

[57] Cf. BOOLE, 1998, p. 37.

[58] É só multiplicar as fórmulas "$y(1-x)=0$" e "$0=zy$", eliminando o termo "y".

[59] Boole não pode usar também $y=y.x$, pois não seria possível eliminar y, usando-se E2.

"$x=x.y$". Além disso, seria possível provar deduções do tipo: "Todo X é Y" implica "Algum X é Y".

1. $x=v.y$	Premissa
2. $v=v$	Identidade
3. $v.x=v^2y$	1,2 **RFM**
4. $v^2=v$	Lei Index
5. $vx=vy$	3,4

A interpretação de (5), segundo Boole, em *The Laws of Thought*, é "Algum X é (algum) Y"[60].

O terceiro problema apontado por Corcoran diz respeito ao critério de invalidade sustentado por Boole. Como escrevi anteriormente, a fim de mostrar a invalidade de um argumento, Aristóteles apresenta um contraexemplo. Boole parece não ter notado a simplicidade e naturalidade deste método. Seu critério para estabelecer a invalidade de um argumento é extremamente complexo e não-natural. Assumamos o exemplo de Boole (1998, p. 35):

De

$xy=x$ (Todo X é Y)

e

$z=zy$ (Todo Z é Y)

---

[60] A classe indeterminada foi alvo de muitos debates e que seu uso gera incongruências no sistema de Boole. Por exemplo, Peirce (1989, p. 171) mostrou que da fórmula "$v.x=v.(1-y)$" (Alguns Xs são não-Y) é possível derivar a fórmula "$v.y=v.(1-x)$" (Alguns Ys são não-Xs), o que é falacioso (de *algum mamífero não é humano* não se segue que *algum humano não é mamífero*). A derivação seria da seguinte forma: assuma a fórmula "$vx=v(1-y)$". Ora, multiplicando o segundo termo da equação, obtemos

$v.x=v-v.y$;

passando "vx" para direita e "vy" para esquerda (da identidade), obtemos

$v.y=v-v.x$;

daí obtemos

$v.y=v.(1-x)$,

ou seja, "Alguns Y são não-Xs". Contudo, essa falácia é barrada no sistema de Boole, uma vez que ele assume um conjunto enumerável (?) de símbolos: v, v', v'', etc. Já mencionamos isso na seção 2. De fato, Boole (2003, p. 231) assume um procedimento semelhante na derivação de Barbara. Se assumirmos esse conjunto enumerável (?), então "algum X não é Y" poderia ser traduzida por $v.x=v'.(1-y)$ e "algum Y não é X", por $v''.y=v''.(1-x)$.

é obtido, usando-se **E2**,

$$xz=xz;$$

para Boole, isto é o mesmo que

$$0=0.$$

Segundo Boole, todo argumento que produz esta equação é inválido.- Porém, esta não é a única maneira de se mostrar a invalidade de um argumento. Em *The Mathematical Analysis of Logic*, Boole assume certas equações auxiliares para mostrar a dedução de "0=0" das premissas. Ele escreve:

> Of cases in which no inference is possible, we take as examples – AO, Fig. 1, and by mutation, OA, Fig. 4,
>
> All Ys are Xs $y(1-x)=0$, $y(1-x)=0$
>
> Some Zs are not Ys $vz=v(1-y)$ (*a*) $v(1-z)=vy$
>
> $v(1-z)(1-x)=0$ (*b*)
>
> $0=0$
>
> since the auxiliary equations in this case is $v(1-z)=0$ (BOOLE, 1998, p. 38).

De acordo com Corcoran, estas equações auxiliares estão diretamente ligadas com a interpretação de "*v*" como uma espécie de quantificador. Se, nos contextos nos quais "*v*" ocorre, ele significar "Alguns Xs", então a equação auxiliar será "$v(1-x)=0$"[61]. Se ele significar "Alguns não-Xs" a equação auxiliar será "$vx=0$". No caso acima, "*vz*", em "$vz=v(1-y)$", significa "Alguns Zs", daí a necessidade da equação auxiliar "$v(1-z)=0$", para se obter a equação "0=0".

[61] Embora confuso, isto tem alguma coerência. Se "*v*" no contexto "*vx*" significar "Algum X", então a classe "*v*" é composta apenas de elementos que pertencem à classe *x*. Logo, a interseção entre a classe *v* e a classe não-*x* tem de ser vazia.

5. Conclusão

George Boole, ao publicar o livro *The mathematical analysis of logic*, estabeleceu o primeiro estágio da evolução da lógica matemática moderna. Podemos listar alguns de seus feitos: (1) criação de uma linguagem algébrica na qual é possível (2) expressar leis lógicas, (3) métodos de inferência que preservam verdade (excetuando-se seu método de solução de equações). Além disso, (4) Boole foi capaz de interpretar as sentenças categóricas da silogística aristotélica na sua linguagem e (5) derivar as figuras silogísticas a partir dos axiomas lógicos.

Na introdução, mencionei a posição defendida por Chateaubriand em relação a Boole, segunda a qual o lógico britânico não pode ser considerado o pai da lógica matemática moderna, pois essa nasceu a partir da concepção metafísica de lógica de Frege e Russell (lógica como ciência). De fato, tanto em *The mathematical analysis of logic* como em *The laws of thought* Boole sustentou que os princípios lógicos não são metafísicos, mas matemáticos. Entretanto, como tentei mostrar na seção 3, essa posição de Boole trata-se de um erro, que levou seu sistema de lógica a um caminho tortuoso, impedindo-o de aceitar leis lógicas bastante naturais como $1+x=1$ e $x+x=x$. Não obstante, Boole defende que lógica é uma ciência que lida tanto com relação entre classes como com a relação entre proposições. Na minha visão, a separação entre metafísica e lógica defendida por Boole é uma mera ilusão.

Apesar dos grandes feitos do sistema de Boole, tentei mostrar também que o seu sistema lógico tem deficiências. Uma delas já mencionei acima: a não-aceitação de axiomas lógicos naturais. Mas há também o problema das soluções de equações eletivas, que está atrelada à conversão parcial. De acordo com a opinião de Corcoran, com a qual concordo, a solução de equações não é um método de dedução. Argumentei que as fórmulas (1) $x=vy$ e (2) $x(1-y)$ que seriam supostamente equivalentes não têm o mesmo conjunto de deduções. Portanto, isso coloca em xeque a derivação de (1) a partir de (2), usando-se o método de solução. Como mostrei, a derivação de (2) a partir de (1) faz uso apenas dos axiomas do sistema e dispensa o método de solução. Finalmente, mencionei rapidamente o método de refutação de Boole.

Em comparação com método de contraexemplos de Aristóteles, o método de Boole é complexo e não-natural.

Referências

ARISTÓTELES. *The Works of Aristotle* vol. I. Publicado por William Benton. Chicago: Encyclopaedia Britannica, 1952.

BOOLE, G. George Boole: *Selected Manuscript on Logic and its Philosophy*. Editado por Ivor Grattan-Guinness e Gérad Bornet. Berlin: Birhäuser Verlag, 1997.

BOOLE, G. *The Mathematical Analysis of Logic.* England: Thoemmes Press, 1998.

BOOLE, G. The Laws of Thought. New York: Prometheus Books, 2003.

CHATEAUBRIAND, O. Boole on Reference and Universe of Discourse: Reply to John Corcoran. IN: SILVA, J. J. da (Ed.). *Logic, Truth and Description*: Essays on Chateaubriand's Logical Forms. Campinas, São Paulo: Manuscrito, 2004. p. 173–182.

CORCORAN, J. Boole's Criteria for Validity and Invalidity. *Notre Dame Journal of Formal Logic,* v. 21, n. 4, p. 609–638, 1980.

CORCORAN, J. Aristotle's Prior Analytics and Boole's Laws of Thought. *History and Philosophy of Logic*, v. 24, p. 261–288, 2003.

CORCORAN, J. The Principle of Wholistic Reference. IN: SILVA, J. J. da (Ed.). *Logic, Truth and Description*: Essays on Chateaubriand's Logical Forms. Campinas, São Paulo: Manuscrito, 2004. p. 159–172.

DUMMETT, M. George Boole. IN: DUMMETT, M. (Ed.). *Truth and Other Enigmas*. London: Duckworth, 1978. p. 66–73.

HEIJENOORT, J. van (Ed.). *From Frege to Gödel*. Cambridge: Harvard University Press, 1981.

LEWIS, C. I.; LANGFORD, C. H. *Symbolic Logic*. New York: Dover Publications, 1959.

PEIRCE, C. S. On the algebra of logic. IN: KLOESEL, C. J. W. et. all (eds.) *Writings of Charles S. Peirce*: a chronological edition, v. 4. Bloomington e Indianapolis: Indiana University Press, 1989, p. 163-209.

QUINE, W. van O. *Mathematical Logic*. New York: Harper and How, 1940.

SENSE, REFERENCE AND COMPUTATION

Bruno Bentzen[1]

ABSTRACT

In this paper, I revisit Frege's theory of sense and reference in the constructive setting of the meaning explanations of type theory, extending and sharpening a program–value analysis of sense and reference proposed by Martin-Löf building on previous work of Dummett. I propose a computational identity criterion for senses and argue that it validates what I see as the most plausible interpretation of Frege's equipollence principle for both sentences and singular terms. Before doing so, I examine Frege's implementation of his theory of sense and reference in the logical framework of *Grundgesetze*, his doctrine of truth values, and views on sameness of sense as equipollence of assertions.

Keywords: Sense. Reference. Computation. Type theory.

RESUMO

Neste artigo, eu revisito a teoria do sentido e referência de Frege no cenário construtivo das explicações de significado da teoria dos tipos, estendendo e aprimorando a análise em termos de programa-valor da teoria de sentido e referência proposta por Martin-Löf, baseada em trabalhos anteriores de Dummett. Eu proponho um critério de identidade computacional para sentidos e defendo que ele valida o que considero a interpretação mais plausível do princípio de equipolência de Frege, tanto para sentenças quanto para termos singulares. Antes de fazer isso, examino a implementação de Frege de sua teoria dos sentidos e referência no quadro lógico de *Grundgesetze*, sua doutrina dos valores da verdade, e sua interpretação de igualdade de sentido como equipolência de afirmações.

Palavras-chave: Sentido. Referência. Computação. Teoria dos tipos.

[1] Pós-doutorando na Carnegie Mellon University – CMU.
Email: b.bentzen@hotmail.com ORCID: 0000-0002-5987-7806

1. Introduction

Frege was one of the first to dispute that all mathematical truths are based on intuition, the cornerstone of the dominant Kantian philosophy of mathematics, and put forward the thesis that all arithmetical truths are reducible to logical truths.[2] The success of his logicism would eventually depend on the plausibility of the development of a new system of logic in which all the objects and concepts of arithmetic could be defined. Indeed, that was the original purpose of his *Begriffsschrift*, the book which lays down the first systematic treatment of modern logic, the ideography, and applies the system to the mathematical theory of sequences. Frege hints at his logicist program in the preface to his book, announcing that his investigations will be continued with the logical elucidation of the concepts of number, magnitude, and so forth in what he describes as an immediately following publication. Yet, it would not be until fourteen years later that the promised book would finally appear in the form of the first volume of *Grundgesetze*. Frege himself explains that this long delay was due to the accommodation of essentially two major changes to the ideography, a task that ultimately forced him to discard an almost completed manuscript.[3] The first change was the introduction of value-ranges of functions, new objects that are intended to generalize the informal notion of extension of concept that he invokes in §68 of *Grundlagen* to overcome the infamous Julius Caesar problem with an explicit definition of the concept of number. This point has been studied extensively and it will not concern us here.[4] Instead, I would like to focus on Frege's theory of sense and reference, the doctrine that singular terms and complete sentences have a mode of presentation and denote an object, whose implementation consists in the second substantial change made to the ideography.

Unfortunately, Frege's formal account of his distinction between sense and reference leaves much to be desired, as he failed to develop his ideography sufficiently in order to capture his answer to the paradox of identity,

[2] References to Frege's works will be cited by name only and indicated by page or section number in the text. I shall ambiguously refer to the formal language and formal theory of *Grundgesetze* by "ideography", unless explicitly stated otherwise

[3] Cf. *Grundgesetze* I, ix.

[4] For a recent survey on the topic, see Bentzen (2019a).

namely, the problem of explaining the apparent difference in cognitive value between true identity statements of the form $a = a$ and $a = b$. This will be the main subject of Section 2, where I shall examine the role of the theory of sense and reference in the technical development of Frege's logicist project and his views on sameness of sense as equipollence of assertions. In Section 3, I shall shift gears and turn to an opposing philosophical position that I take to possess a more suitable semantic setting for the establishment of a proper theory of sense and reference, namely, the constructive conception of mathematics. The idea of reinterpreting Frege's traditional semantic distinction in a constructive setting is certainly not new, as it has been first explored mainly by Dummett (1978) and Martin-Löf (2001). The latter author has extended considerably the ideas of the first one by reshaping them in the context of his type theory, a formal system in which every term is assigned to a type and every operation strictly restricted to terms of a certain type.[5] But at the same time several aspects of Martin-Löf's account strike me as incomplete or problematic, especially because he does not seem to make full use of the powerful intended semantics of type theory, an informal realizability interpretation known as the meaning explanations (Martin-Löf, 1982). My goal in this paper is to revise some of Martin-Löf's main views, a task that will be carried out in Section 4, where I carry the meaning explanations to their ultimate logical conclusions.

2. Frege on sameness of sense

In Frege's works we find the two forms of judgment that are current in modern logic, the judgment form that asserts that a proposition is true, which is written in turnstile notation as

$$\vdash A$$

[5] [4] In this paper, I will be concerned with a fragment of Martin-Löf (1975) where the only types are dependent function types, equality types, and the natural numbers. Judgments of typehood are used instead of type universes, and typehood and type membership judgments are defined via primitive type and term judgments, as described in Section 3.

and the judgment form that states that a sentence expresses a proposition, which is implicitly introduced for the first time in §2 of *Begriffsschrift* via the content stroke

$$\text{—— } A$$

Therein, a proposition is described as a "judgeable content" and identified with the content of a turnstile judgment. This intuitive notion of content represents the main semantic element of the ideography at this earlier point and, in fact, logical operations such as implication, negation, and universal quantification are all exclusively applicable to contents, or, more precisely, to judgeable contents or propositions.

The only exception to this structure is the relation of identity of content which is introduced in §8 of *Begriffsschrift* as a metalinguistic relation between the expressions themselves and not their respective contents. That is, according to this interpretation of equality, an equality statement $a = b$ means that the expressions '*a*' and '*b*' have the same content. One semantic consequence of this is that two propositions cannot be identified if they imply each other, for they may have different contents. Indeed, the propositional extensionality principle

$$(A \supset B) \supset (B \supset A) \supset (A=B)$$

is not only not a theorem in *Begriffsschrift* but also inconsistent with its formal system, as noted in Duarte (2009). However, once the ambiguous notion of content is split into sense and reference, propositional extensionality becomes a valid principle for references, although not for senses, considering that from the fact that two theorems are logically equivalent it does not follow that they must have the same cognitive value.[6]

[6] Curiously, Frege (1982) once considered an identity criterion for senses along those lines in a letter to Husserl dated 9 December 1906, but this suggestion is clearly unacceptable, since it would mean that every two true propositions $a = a$ would $a = b$ express the same sense.

2.1. The doctrine of truth values

Frege's theory of sense and reference is systematically introduced for both singular terms and complete sentences in his seminal essay *Sinn und Bedeutung*. The theory seems to provide a compelling answer to the paradox of identity with respect to singular terms, for we are able to say that the cognitive value of $a = a$ and $a = b$ is generally different because a and b may differ in their sense, despite being coreferential. However, the theory becomes highly controversial when we consider the sense and reference of sentences, which are taken to express a thought (in modern terminology, a proposition) and refer to a truth value, which, since referents are always objects, are seen as the simplest representatives of logical objects. Frege is well aware of this counterintuitive aspect of his doctrine, as he immediately attempts to justify it by noting that what he calls an object can only be properly understood in connection to concepts and relations and then attempts to substantiate his claim that the reference of a sentence is its truth value by remarking that the latter remain unchanged when a part of the sentence is replaced by a coreferential expression, that is, that sentences with the same truth value are interchangeable *salva veritate*. Later on, in the preface of *Grundgesetze*, Frege remarks that the introduction of truth-values may seem strange at first, but adds that the fact that everything becomes much simpler and sharper with those objects puts a great weight in the balance in favor of his own conception.[7] In the ideography, concepts and relations are treated as particular cases of functions that output truth values and Frege explicitly states that an object is anything that is not a function.

Besides the poor justifications for Frege's adoption of his strange doctrine of truth values, it is disappointing that almost no considerations about sense or thoughts can be found in *Grundgesetze*, at least if we assume that the real purpose of the distinction between sense and reference is to offer a solution to the paradox of identity. Although Frege still seems to have a strong conviction that a criterion of identity for senses is of fundamental importance, as

[7] [6] Cf. *Grundgesetze* I, x.

we can judge from, for instance, a December 1906 letter to Husserl, where Frege writes:

> It seems to me that an objective criterion is necessary for recognizing a thought again as the same, for without it logical analysis is impossible. (FREGE, 1982, p. 70)

Frege does not seem to bother to provide a logic of sense and reference in the book he intends to show once and for all that arithmetic is nothing but logic. For this reason, we can say that Frege failed to show that some mathematical equality statements have cognitive value in *Grundgesetze*, because his solution to the identity paradox is entirely dependent on what it means for two expressions to have different senses, and no identity criterion for senses is given in his ideography.

It is possible that Frege had come to recognize that the most important notion for the vindication of his logicism is that of reference, as pointed out by Simons (1992), and that for those purposes all that we need to know about sense is that a sentence needs to express a thought in order to refer at all. In fact, now the horizontal stroke symbolism

$$\text{——}\ A$$

which used to indicate that *A* is a proposition (judgeable content) before the division of content into sense and reference, no longer says that *A* expresses a proposition (thought). Instead of being an implicit form of judgment, the horizontal is treated as a function that refers to a truth value depending on the reference of *A*, yielding the true if *A* denotes the true and the false otherwise. Similarly, negation does not operate on contents anymore, for it is explicitly treated as a function

$$\text{—}\vdash A$$

that has the opposite effect of yielding the false if *A* does not refer to the true and the true otherwise. The turnstile judgment has kept its role, being the only judgment form of the revised ideography. It is now fully explained in terms of reference as well, now asserting that the expression *A* refers to the true. There is no way to formally assert the fact that a sentence expresses a thought in *Grundgesetze*, which is to say to fulfill the role previously played by the content stroke.

In general, it would seem that Frege contradicts himself by downplaying the value of sense and treating his doctrine of truth values as a mere technical device for the successful accomplishment of his logicism, as argued in Ruffino (1997) and Duarte (2009). This can be seen more clearly in light of the discarded manuscript that was written before the introduction of the distinction between sense and reference to the ideography in *Grundgesetze*. In §69 of *Grundlagen*, immediately after the explicit definition of the concept of number is proposed by Frege, he attempts to confirm the fruitfulness of his definition by sketching a proof of a crucial theorem that says that the number that falls under the concept *F* is equal to the number that falls under the concept *G* iff *F* and *G* are in one-to-one correspondence (Hume's Principle). The proof consists in showing that both sides of the biconditional imply each other. But in the ideography, where a biconditional is always represented by an equality, the formalization of the proof cannot be completed without propositional extensionality. In *Grundgesetze*, what makes the derivation of propositional extensionality as a theorem possible is the introduction of Basic Law IV, which roughly states that the truth values denoted by the sentences *A* and *B* either coincide or not

$$\neg(\text{——}A = \text{—}\vdash\text{—}B) \supset (\text{——}A = \text{——}B)$$

But this is only allowed as an axiom in the ideography because of the distinction between sense and reference. If those observations are correct, Frege conceived his theory of sense and reference primarily as a logical apparatus necessary to overcome technical obstacles rather than as a solution to the

paradox of identity as it is commonly thought.[8] That would explain why Frege does not do justice to the notion of sense in *Grundgesetze*.

2.2. The equipollence principle

What comes closer to the proposal of a criterion of identity for senses in Frege's writings is the equipollence principle that two sentences *A* and *B* express the same thought provided that anyone who accepts *A* as true must also immediately accept *B* as true and vice versa. Sundholm (1994) has called attention to the fact that this idea has been informally suggested on many occasions by Frege, most notably in the opening passage of his *Kurze Übersicht meiner logischen Lehren*:

> Now two propositions *A* and *B* can stand in such a relation that anyone who recognizes the content of *A* as true must thereby also recognize the content of *B* as true and, conversely [...] So one has to separate off from the content of a proposition the part that alone can be accepted as true or rejected as false. I call this part the thought expressed by the proposition. (FREGE, 1906, pp.197–98)

I take the following equipollence to be the most natural choice of a corresponding principle for singular terms: two singular terms *a* and *b* express the same sense if for every predicate *P* , anyone who accepts *P* (*a*) as true must also immediately accept *P* (*b*) as true and vice versa. Related to *salva veritate*, this idea is already hinted at in *Sinn und Bedeutung*

> If we now replace one word of the sentence by another having the same reference, but a different sense, this can have no bearing upon the reference of the sentence. Yet we can see that in such a case the thought changes; since, e.g., the thought in the sentence 'The morning star is a body illuminated by the Sun' differs from that in the sentence 'The evening star is a body illuminated by the Sun.' Anybody who did not know that the evening star is the morning star might hold the one thought to be true, the other false. (FREGE, 1892, p.62)

Yet, much remains to be done before both equipollence principles for singular terms and sentences can be adopted as satisfactory identity criteria for senses, for we do not have a rigorous account of what it means to "immediately" accept a proposition as true. In *Funktion und Begriff*, Frege has ex-

[8] [7] See e.g. Ruffino (1997, §3) and Duarte (2009, p.167).

pressly stated that the sense of two expressions is equal up to renaming of bound variables, that is, that we could write $x^2 - 4x$ directly as $y^2 - 4y$ without altering its sense, however, it can be very difficult to determine the extent of strictness in his account of sameness of sense, considering that Frege also happens to claim that the two halves of Basic Law V express the same sense, but in a different way.[9]

3. Constructive semantics in type theory

In the remainder of this paper, I will argue that a constructive rendering of the theory of sense and reference is capable of not only providing a precise computational interpretation to the equipollence principles for singular terms and sentences, but also validating them. I will begin our discussion with a brief outline of some important aspects of constructive semantics.

In Frege's semantics, as mentioned in the previous section, the assertion that a proposition is true is understood as declaring the fact that the thought expressed by a sentence denotes the true, an interpretation that rests on a realist assumption that numbers and other sorts of mathematical objects are non-physical and non-mental entities that do not exist in space or time. Therefore, the most fundamental form of judgment in logic and mathematics is the truth judgment, which Frege writes as

$$\vdash\!\!\!— A$$

In contrast, the constructivist tradition states that the only legitimate way of understanding a proposition is as a specification of a construction with certain given properties and, moreover, that in order to assert that a proposition is true one has to exhibit a construction that realizes the specification expressed by the given proposition (Martin-Löf, 1985). The central form of

[9] Both claims are found in Frege (1891, p.27). Frege's views on sameness of sense have been dealt with in more detail in Sundholm (1994, p.304–307), Klement (2016, §4), and Bentzen (2020c, §2.1).

judgment is therefore the one that states that *a* is a construction that realizes a proposition *A*, which, using the language of type theory, is often expressed as

$$a : A$$

Put differently, to assert that a proposition is realized by a construction is to declare the truth of that proposition. One thing that should be emphasized is that the existence of a construction should not be understood as an ontological existential claim that may be verified by means independent of our knowledge, but rather as the epistemic act of conceiving such a mathematical construction. Accordingly, the constructive conception of truth can be regarded as a form of anti-realism with respect to mathematical objects.

3.1. Judgments

In type theory, a construction is a humanly computable procedure or program that is generally accepted as a method for obtaining a mathematical object. Following the so-called propositions-as-types correspondence (Howard, 1980), both propositions and sets are specifications of constructions that are structured through the unifying concept of a type. The elements of a type are commonly called terms, and, as expected, terms are interpreted as constructions.

For a more concise semantic explanation, it is convenient to base our type theory on two equality forms of judgment that state that two expressions are equal types and also that two expressions are equal terms of a type. I shall write those equalities with a triple bar notation to emphasize the fact that they are judgmental relations, meaning that they only occur in assertions and therefore are not subject to operations such as those determined by logical connectives

$$A \equiv B \text{ type} \qquad\qquad a \equiv b : A$$

The reflexivity of those equality judgments can be used to define their more common counterparts A type and $a : A$, which say that an expression is a type and that an expression is a term of a type, as $A \equiv A$ type and $a \equiv a : A$. In other words, to determine their meaning, it suffices to determine the meaning of the type and term equality as primitive judgments. The general strategy is that the meaning of a judgment is given via an untyped model of computation, an idea that derives directly from the meaning explanations (Martin-Löf, 1982). Every meaningful statement is based on a notion of computation that is accepted as primitive and used to give terms a computational behavior, define types as term specifications, and to assign terms to types based on the values to which they compute.

3.2. Computation

The whole process of meaning explanation starts with the specification of a native computation system that takes the form of a programming language, typically the untyped lambda-calculus extended with constants for the type formers, constructors, and eliminators of the type theory. In this paper, I will consider dependent function types $\Pi_{x:A}B$, equality types $a =_A b$, and the natural numbers *nat*. First, we specify the syntax of the programming language

$$
\begin{aligned}
&\mathsf{var} \;:= x \mid y \mid z \mid \ldots \mid x' \mid y' \mid z' \mid \ldots \\
&\mathsf{expr} := \mathsf{var} \\
&\qquad \Pi_{\mathsf{var}:\mathsf{expr}}\mathsf{expr} \mid \lambda \mathsf{var}.\mathsf{expr} \mid \mathsf{expr}(\mathsf{expr}) \\
&\qquad \mathsf{expr} =_{\mathsf{expr}} \mathsf{expr} \mid \mathsf{refl}(\mathsf{expr}) \mid \mathsf{eqrec}(\mathsf{expr}, \mathsf{expr}, \mathsf{expr}) \\
&\qquad \mathsf{nat} \mid 0 \mid \mathsf{succ}(\mathsf{expr}) \mid \mathsf{natrec}(\mathsf{expr}, \mathsf{expr}, \mathsf{expr})
\end{aligned}
$$

Then, we endow this language with an operational semantics, a complete description of how terms are expected to compute given in the form of a transition relation over closed terms, that is, expressions with no occurrence of free variables, and reflexive on certain closed terms, which are regarded as execution values.

This transition relation is fully captured by the computation rules that I now symbolically describe, where

$$a \longmapsto a'$$

indicates the fact that a transitions to a' and a val means that a is a value, which is to say that a evaluates to a. I will start considering the rules for the dependent function type, which generalizes the notion of universal quantification and indexed product of sets

$$\frac{}{\prod_{x:A} B \text{ val}} \qquad \frac{}{\lambda x.a \text{ val}}$$

$$\frac{a \longmapsto a'}{a(b) \longmapsto a'(b)}$$

$$\frac{}{(\lambda x.a)(b) \longmapsto a[b/x]} \qquad \frac{}{\lambda x.a(x) \longmapsto a}$$

The dependent function type $\Pi_{(x:A)}B$ is inhabited by functions $\lambda x.a$, where a is an open term that may depend on x, and if f is a function and a a term, then $f(a)$ is the application of f to a. The last two computation rules are of particular importance, because they induce some obvious resemblances between lambda terms $\lambda x.f(x)$ and Frege's value-ranges $\grave{\varepsilon}f(\varepsilon)$, since $\grave{\varepsilon}f(\varepsilon) \cap a = f(a)$ is a theorem in the ideography (see *Grundgesetze* §34), although $\grave{\varepsilon}f(\varepsilon) = f$ contradicts the intended semantics of the ideography since value-ranges are taken to be objects and cannot be functions. I explore this connection in detail in Bentzen (2020b), which contains a more comprehensive type-theoretic study of the ideography.

The following rules provide a full computational account of the equality type $a =_A b$, the type-theoretic counterpart of the usual notion of an equality proposition. Unlike the judgmental equality $a \equiv b : A$, $a =_A b$ does not have any assertive force. The constructor refl(a) represents the reflexivity of equality and the eliminator eqrec(a, b) generalizes the principle of *salva veritate* that Frege borrows from Leibniz.

$$\frac{}{a =_A b\ \mathsf{val}} \qquad \frac{}{\mathsf{refl}(a)\ \mathsf{val}}$$

$$\frac{a \longmapsto a'}{\mathsf{eqrec}(a,b) \longmapsto \mathsf{eqrec}(a',b)}$$

$$\frac{}{\mathsf{eqrec}(\mathsf{refl}(a),b) \longmapsto b}$$

Finally, we have the rules for natural numbers, which give us 0, a successor operator succ(*n*) and an explicit recursor natrec*(a, b, c)* that incorporates the principle of complete mathematical induction

$$\frac{}{\mathsf{nat}\ \mathsf{val}} \qquad \frac{}{0\ \mathsf{val}} \qquad \frac{}{\mathsf{succ}(a)\ \mathsf{val}}$$

$$\frac{a \longmapsto a'}{\mathsf{natrec}(a,b,c) \longmapsto \mathsf{natrec}(a',b,c)}$$

$$\frac{}{\mathsf{natrec}(0,a,b) \longmapsto a}$$

$$\frac{}{\mathsf{natrec}(\mathsf{succ}(a),b,c) \longmapsto (c(a))(\mathsf{natrec}(a,b,c))}$$

As it can be seen, the basic pattern here is that all type formers and their constructors are interpreted as values, eliminators preserve computation on their main argument, and eliminators transition to a particular information that was given to them for when their main argument is a certain constructor.

The transition relation divides terms into two classes: those that are themselves values are called canonical, and those that are not themselves values but may eventually reach a value if we keep iterating the transition process are called non-canonical. This iteration is made precise with the notion of evaluation, a computation that halts when it finds the value that a term transitions to after an indefinite number of transition steps. I shall write *a ↓ a'* to mean that *a* evaluates to a'. Since we are dealing with untyped computations, the evaluation of a term will not always terminate, but it suffices to characterize closed expressions as programs that when executed output the value they evaluate to. Notice that natural numbers are considered as a canonical terms whereas, taking into account Frege's conception of numbers as extensions of

concepts, one might expect such number terms to evaluate to lambda terms, the canonical terms representing value-ranges. This can be done with a technique known as Church encoding, with which we can represent the natural numbers using lambda notation. But the idea goes against the spirit of type theory in that it conflates two very distinct data types, numbers and functions.

3.3. Type and term equality

Roughly, we explain what a type is by specifying the terms of that type following an idea that can be traced back to the constructive conception of set advocated by Bishop (1967) and the interpretation of the intuitionistic logical constants proposed by Heyting (1934). In fact, the first thing we do to explain what types are is distinguishing between canonical and non-canonical types. Any expression that evaluates to a canonical type is a type, and a canonical type is explained by prescribing what their canonical terms are. The meanings of the type and term equality judgments are mutually explained in a similar way. To account for the former, we assume that A and A' are closed expressions, and assert that

> *A and A' are equal types provided that they evaluate to the same canonical type up to type equality.*

This characterizes type equality as a relation between terms that behave in a certain expected way. The following rules illustrate how type equality can be explained for our canonical types: dependent function types, equality types, and natural number types

$$\frac{A \downarrow \prod_{x:B} C \quad A' \downarrow \prod_{x:B'} C' \quad B \equiv B' \,\mathsf{type} \quad x : B \vdash C \equiv C'}{A \equiv A' \,\mathsf{type}}$$

$$\frac{A \downarrow a =_B b \quad A' \downarrow a' =_{B'} b' \quad B \equiv B' \,\mathsf{type} \quad a \equiv a' : B \quad b \equiv b' : B}{A \equiv A' \,\mathsf{type}}$$

$$\frac{A \downarrow \mathsf{nat} \quad A' \downarrow \mathsf{nat}}{A \equiv A' \,\mathsf{type}}$$

While the above stipulations endow type equality with an intensional nature (Dybjer, 2012), we often find an additional condition that determines a type uniquely by its terms

$$\frac{a : A \vdash a : A' \quad a : A' \vdash a : A}{A \equiv A' \ \mathsf{type}}$$

Now, in order to give a full account of term equality we assume that we are given closed terms a and b as well as a type A, and evaluate them all, declaring that

> *a and b are equal terms of type A provided that a and b evaluate to equal canonical terms of the canonical type which A evaluates to.*

More concretely, term equality can be explained by laying down what it is to evaluate to equal canonical terms up to term equality for all the considered canonical types:

$$\frac{A \downarrow \prod_{x:B} C \quad a \downarrow \lambda x.M \quad a' \downarrow \lambda x.M' \quad x : B \vdash M \equiv M' : C}{a \equiv a' : A}$$

$$\frac{A \downarrow b =_B b' \quad a \downarrow \mathsf{refl}(b) \quad a' \downarrow \mathsf{refl}(b') \quad b \equiv b' : B}{a \equiv a' : A}$$

$$\frac{A \downarrow \mathsf{nat} \quad a \downarrow 0 \quad a' \downarrow 0}{a \equiv a' : A}$$

$$\frac{A \downarrow \mathsf{nat} \quad a \downarrow \mathsf{succ}(M) \quad a' \downarrow \mathsf{succ}(M') \quad M \equiv M' : \mathsf{nat}}{a \equiv a' : A}$$

The semantics given above can be inductively extended to judgments occurring under a list of hypothesis, meaning that the validity of the hypothetical judgment

$$x : A \vdash f : B$$

is subject to the validity of the non-hypothetical or categorical judgment $f[a/x] : B$ for every closed term $a : A$. In sum, the only way a hypothetical judgment can be given meaning to is by determining what categorical judgment they result in when all their free variables are replaced with closed terms. Frege's determination of the reference of functional expressions in *Grundgesetze* §§29–31 is based on a similar idea, since the role that functional expressions play in the ideography is no different than that of open terms in type theory (Bentzen, 2020b).

Because only closed terms are taken into account in the validation of a hypothetical judgment, it can be shown that the following principle, which can be regarded as a stronger version of the so-called Axiom K (Streicher, 1993) obtains

$$x : A, p : x =_A x \vdash p \equiv \mathrm{refl}\ (x) : x =_A x$$

because, for a closed term $a : A$, a closed equality term $b : a =_A a$ will always evaluate to a canonical term of $a =_A a$, which means that $p \downarrow \mathrm{refl}(x)$. Since $a =_A a$ is a canonical type, $c \equiv \mathrm{refl}(a) : a =_A a$ obtains. This principle is a straightforward consequence of the reflection rule

$$\frac{p : x =_A y}{x \equiv y : A}$$

which can be easily validated using a similar line of reasoning and expresses the fact that, semantically, every propositional equality is also a judgmental equality.

4. Sense and reference

It was Dummett (1978) who first noticed that Frege's theory of sense and reference could be rendered constructively by assuming that the sense of a singular term is related to its reference as a program is related to its value, and that the thesis that the reference of a sentence is a truth value should be enti-

rely rejected, since, if we think of the sense of an expression not as a mode of presentation of its reference but as an effective method for determining it, that would imply a means of deciding whether a proposition is true or false.[10]

Although the idea that a sentence refers to a truth value has no place in a constructive setting, Martin-Löf (2001) has observed that an interpretation of singular terms as terms and propositions as types in the sense of the meaning explanations allows for a more rigorous development of a theory of sense and reference for both singular terms and sentences. In this context, there is nothing better suited to mediate the passage from the sense to the reference of an expression than the evaluation of a term to its canonical form. More concretely, the reference of a term $(\lambda x.\mathrm{refl}(x))(0)$ is the canonical term it evaluates to, refl(0). Programs are means for specifying their values and so are senses, since values are references. Expanding this idea to the sphere of types, Martin-Löf notes that the reference of a sentence has to be a canonical proposition, a view that fits perfectly the meaning explanations, since an expression that is assigned to a type is interpreted as a sentence that expresses a proposition, and a type evaluates to a canonical type, which is taken as their referent. For example, both $(\lambda x.x)(\mathrm{nat})$ and $\mathrm{natrec}(0, \mathrm{nat}, \lambda x.x(x))$ are equal types because they evaluate to the canonical type nat. In other words, they are coreferential types but specify their value differently.

If the passage from the sense to the reference of an expression is given through evaluation, how should one interpret the passage of an expression to its sense? I have suggested elsewhere that an expression a comes to be known as a program when it is assigned to a type A.[11] The reason for this lies in an analogy with computer programs: a piece of code is only recognized as a program when it is correctly typechecked by a compiler of the programming language in question. If there is a single type mismatch error, such as an assignment of a value between two variable of different types, then we did not write a program after all. However, if we assume that a term a has a sense only when $a : A$, we are at the same time excluding the possibility that terms

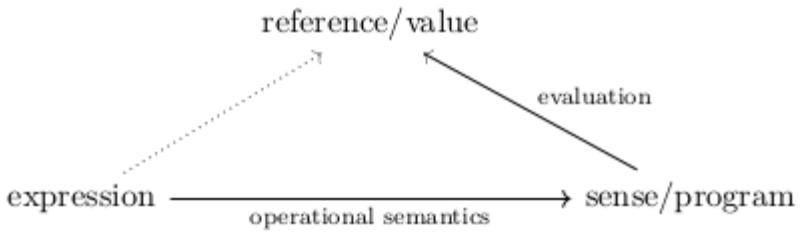
[10] Moschovakis (1994) adds to this interpretation a mathematical notion of recursive algorithm that allows for a more rigorous theory of sense. It can be seen as a forerunner of the theory of sense of reference of Martin-Löf (2001), which is grounded in type theory instead.

[11] See, for instance, Bentzen (2018, 2020c).

lacking a reference have a sense. This has to do with the fact that $a : A$ obtains only when a computes to a value according to our semantic stipulations, while not every term has a value. Consider a non-terminating term such as the infamous Ω term

$$(\lambda x.x(x))(\lambda x.x(x))$$

Frege has always defended that in an exact science such as mathematics every expression must have a reference, but he never once entertained the idea that a sentence does not express a sense if it lacks a reference. Even in our setting, it is desirable to allow expressions to have a sense when they do not have a reference, for even a non-terminating term can be run just like any other program. I shall therefore say that an expression is recognizable as a program when it is interpretable by an operational semantics in the style of the previous section. That means that all that is needed for an expression to have a sense is a grammatical structure that conforms to the program syntax. Returning to our analogy with computer programs, we may say that a piece of code is a legitimate program when it passes a lexical check, when there are no missing required characters or unrecognizable tokens. Using Frege's semantic triangle, the tripartite distinction between expressions, programs, and values can be succinctly formulated as follows



Martin-Löf is careful to distinguish between two different interpretations of equality of reference, claiming that a judgmental equality $a \equiv b : A$ says of the senses of a and b that they are coreferential and a propositional equality $a =_A b$, when proven to be true, says of the references of the expressions 'a' and 'b' that they are equal objects, because one is here suppressing the equality term p that realizes the propositional equality via the assertion of $p : a$

$=_A$ *b*, the part of the judgment that he sees as not "referentially transparent". Having said that, I cannot see the grounds for this confusing distinction, considering that semantically every propositional equality is a judgmental equality. Since Martin-Löf speaks of judgmental equality as an intensional relation, one may question if he is not simply leaving the semantics aside and speaking of the formalism itself, since, to give an example, while it is easy to construct a term of type $n + m =_{nat} m + n$ via the principle of complete induction, there is no derivation of the judgmental equality $n + m \equiv m + n : nat$ in the theory, since both sides of the equality do not evaluate to the same value, although they do for every particular instance of *n* and *m*. Semantically, both propositional and judgmental equalities say that the senses of *a* and *b* have the same reference, because they amount to the fact that the values of programs *a* and *b* are one and the same.

Martin-Löf also claims that equality of sense is given by a relation of synonymy for expressions, which includes in particular renaming of bound variables and definitional stipulations such as $\pi = C/d$. I believe that this view of sameness of sense is essentially correct, but it lacks a theoretical motivation in the setting of the meaning explanations. The following computational explanation is able to provide a more compatible justification. Since to say that two expressions have the same sense is to say that two terms express the same program in type theory, we just have to justify an identity criterion for programs. Clearly, it seems appropriate to identify programs that have the same computational content. Hardly anyone would say that two programs are identical if one involves more computations than other, even though they may have the same execution value. With that in mind, I propose that *a* and *b* express the same program if they have the exact same computational behavior, if to evaluate *a* is to evaluate *b* and vice versa. It is simply natural to identify two terms up to renaming of bound variables and definition unfolding, given that they are evaluated in the same way.

4.1. Equipollence

It only remains to be asked how this account of sameness of sense can be linked to the equipollence principles that we have previously discussed. For the sake of argument, let us first assume that A and B are sentences, which is to say that they evaluate to canonical types. We have to show that A and B have the same sense when anyone who accepts A will immediately accept B and vice versa. One interpretation of the immediacy in the latter condition is that proposed in Sundholm (1994)

$$\frac{a : A}{a : B} \quad \text{and} \quad \frac{b : B}{b : A}$$

that says that any construction that realizes A is also a realizer for B and vice versa. But that does not imply that A and B have the same computational behavior. Instead, it says that they have the same reference, since from this co-extensionality it would follow that $A \equiv B$ *type*. That may seem odd from a Fregean perspective, but recall that sentences do not refer to truth values but to canonical types here.

If we choose to interpret immediacy in a purely computational way instead, then we have that anyone who accepts A will immediately accept B if, without any reference to the notions of transition or evaluation, one is justified in an making assertion of the truth of B from one's assertion of the truth of A, or, equivalently, if from one's construction of a term $a : A$ one can construct a (possibly new) term $b : B$ by means that do not involve the computation. It is plain that sameness of sense is given by equipollence under this view, for this condition describes precisely what it means for two types to have the same computational behavior. Finally, if we are given an open type

$$x : A \vdash P\,(x) \text{ type}$$

then the corresponding principle for terms follows directly from the equipollence of types, because to say that two terms $a : A$ and $b : A$ are equipollent is

to explain how anyone who accepts $P\,(a)$ will immediately accept $P\,(b)$, which we already did. Notice that only terms of a same type can be equipollent, like everything else in type theory. It does not make sense to ask whether, say, Julius Caesar is equipollent to 0.

If we were to interpret equality of reference for types in a strictly Fregean sense, that is, following the idea that two sentences are coreferential if they have the same truth value, then obviously the relation that we would be looking for would be that of logical equivalence, which, type theoretically, translates to the existence of two functions $f : A \to B$ and $g : B \to A$ between the types A and B.[12] But it is worth stressing that actual equality of reference does not follow from logical equivalence, for two types may imply each other without computing to the same canonical type. Put differently, if we had a type universe U, whose terms are smaller types, then the following principle of propositional extensionality

$$(A \to B) \to (B \to A) \to (A =_U B)$$

would be false in the meaning explanations, just as it was in *Begriffsschrift* because A and B may have different contents. In general, we do not even have that $g(f(a)) =_A a$ and $f(g(b)) =_B b$ obtain, for every $a : A$ and $b : B$. The only thing logical equivalence tells us is that if we constructed a term of A we know how to find another term of B and vice versa. It comes with no other guarantees, as Sundholm (1994) notes.

4.2. Equality and homotopy

Over the last few years, homotopy type theory (UFP, 2013) has emerged as a new foundation for mathematics that unites type theory and homotopy theory via a homotopical interpretation of type theory where a type A is

[12] We typically write $A \to B$ as an abbreviation for $\Pi_{x:A}B$ when the type B does not depend on $x : A$.

viewed as a space, a term $a : A$ as a point of the space A, an equality term $p : a =_A b$ as a path from point a to point b in the space A and so on.[13]

One of the main ingredients that reinforces this inter-pretation is the univalence axiom, which offers a formal justification for the common view among mathematicians that two mathematical objects are equal when they are isomorphic. Roughly, the univalence axiom implies that two types are identical just in case they are equivalent in a technical sense.

$$\mathrm{ua} : A \simeq B \to A =_U B$$

And this new sort of identification imposes a weaker understanding of propositional equality where in general a closed term $p : a =_A b$ does not entail $a \equiv b : A$, because univalence introduces a new canonical term $\mathrm{ua}(e)$ to the equality type that is not judgmentally equal to a reflexivity term, assuming that e is an equivalence. Considering that $\mathrm{refl}(a)$ is interpreted as the constant path at the point a, univalence ensures the existence of non-trivial paths.

Yet, this homotopical view of equality goes against the constructive theory of sense and reference that I have expounded, precisely because if propositional equality is to express coreferentiality, then the reference of a term cannot be its value in homotopy type theory. In the presence of non-trivial paths, that is, if we are allowed to construct a closed term $p : a =_A b$ such that

$$p \not\equiv \mathrm{refl}(a),$$

then we have that $a \downarrow a'$ and $b \downarrow b'$ but

$$a' \not\equiv b',$$

meaning that the interpretation of propositional equality as equality of values is lost.

[13] For a philosophical introduction of homotopy type theory, see Ladyman and Presnell (2016) and Bentzen (2020a).

References

BENTZEN, B. *Constructive mathematics and equality*. Tese (Doutorado em Lógica) – Sun Yat-sen University, Institute of Logic and Cognition, Guangzhou, 2018.

BENTZEN, B. "Frege on Referentiality and Julius Caesar in Grundgesetze §10". *Notre Dame Journal of Formal Logic*, v.60, n.4, pp.617-637, 2019. Disponível em: https://projecteuclid.org/euclid.ndjfl/1569830414.

BENTZEN, B. "What types should not be". *Philosophia Mathematica*, v.28, n.1, pp.60-76, 2020a. Disponível em: https://doi.org/10.1093/philmat/nkz014.

BENTZEN, B. "Frege's theory of types". Preprint, 2020b. Disponível em: http://philsci-archive.pitt.edu/17367/.

BENTZEN, B. "On different ways of being equal", *Erkenntnis*. Advance online publication, 2020c. Disponível em: https://doi.org/10.1007/s10670-020-00275-8.

BISHOP, E. *Foundations of Constructive Analysis*. Academic Press, New York, 1967.

DUARTE, A. B. *Lógica e Aritmética na Filosofia da Matemática de Freg***e**. Tese (Doutorado em Filosofia) – PUC-Rio, Departamento de Filosofia, Rio de Janeiro, 2009.

DUMMETT, M. "Frege's distinction between sense and reference". In: *Truth and Other Enigmas*, pp.116-144. Harvard Univ. Press, Cambridge, 1978.

DYBJER, P. "Program testing and the meaning explanations of intuitionistic type theory". In: *Epistemology versus Ontology*, pp.215-241. Springer, 2012.

FREGE, G. "Funktion und Begriff". Hermann Pohle, Jena, 1891. Traduzido em Geach e Black (1980, pp.21-41).

FREGE, G. "Über Sinn und Bedeutung". *Zeitschrift für Philosophie und philosophische Kritik*, 100, pp.25-50, 1892. Traduzido em GEACH & BLACK (1980, pp.56-78).

FREGE, G. "A brief survey of my logical doctrines". H. Hermes, F. Kambartel, and Friedrich Kaulbach (eds and trans.). In: *Frege: Posthumous Writings.* Oxford: Basil Blackwell, 1979, pp.197-202, 1906.

FREGE, G. "Letter to Husserl (1906)". In Gottfried Gabriel, Brian McGuinness, and Hans Kaal (eds), *Philosophical and mathematical correspondence*. University of Chicago Press, 1982. Traduzido por H. Kaal.

GEACH, P. & BLACK, M. *Translations from the Philosophical Writings of Gottlob Frege.* Blackwell, Oxford, 3 ed., 1980.

HEYTING, A. *Mathematische Grundlagenforschung, Intuitionismus, Beweistheorie.* Springer, Berlin, 1934.

HOWARD, W. A. "The formulae-as-types notion of construction". In J. P. Seldin and J. R. Hindley, editores, In: *Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, pp.479-490. Academic Press, London, 1980.

KLEMENT, K. C. "Grundgesetze and the sense/reference distinction". In P. A. Ebert and M. Rossberg, editores, In: *Essays on Frege's Grundgesetze*. Oxford University Press, 2016. Disponível em: https://doi.org/10.1093/oso/9780198712084.003.0006.

LADYMAN, J. & PRESNELL, S. "Does Homotopy Type Theory Provide a Foundation for Mathematics". *The British Journal for the Philosophy of Science*, 69(2), pp.377-420, 2016. Disponível em: https://doi.org/10.1093/bjps/axw006.

MARTIN-LÖF, P. "Constructive mathematics and computer programming". In: *Logic, methodology and philosophy of science, VI (Hannover, 1979), volume 104 of Stud. Logic Found. Math.*, pp.153-175. North-Holland, Amsterdam, 1982. Disponível em: https://doi.org/10.1016/S0049-237X(09)70189-2.

MARTIN-LÖF, P. "On the meanings of the logical constants and the justifications of the logical laws". In: *Atti degli incontri di logica matematica. Scuola di Specializzazione in Logica Matematica*, v. 2, pp.203–281. Universitá di Siena, 1985. Reimpresso em *Nordic Journal of Philosophical Logic*, v.1, n.1, pp.11-60.

MARTIN-LÖF, P. "The sense/reference distinction in constructive semantics". Artigo lido em Leiden (Transcrito por B. Jespersen), 2001.

MOSCHOVAKIS, Y. "Sense and denotation as algorithm and value". In J. Väänänen and J. Oikkonen, editors, Logic Colloquium '90: Proceedings of the logic colloquium, Bristol, vol. 2 of Lecture Notes in Logic, pp. 210-249. Springer-Verlag, Berlin, 1994. Disponível em: https://projecteuclid.org/euclid.lnl/1235423715.

RUFFINO, M. "Wahrheitswerte als Gegenstände und die Unterscheidung zwischen Sinn und Bedeutung". *Kritisches Jahrbuch der Philosophie*, v.2, pp.69–118, 1997.

SIMONS, P. "Why is there so little sense in *Grundgesetze*?" *Mind*, v.101, n.404, pp.753-766, 1992. Disponível em: https://doi.org/10.1093/mind/101.404.753.

STREICHER, T. *Investigations into intensional type theory*. Habilitiation Thesis, Ludwig Maximilian Universität, 1993. Disponível em: http://www.mathematik.tu-darmstadt.de/~streicher/HabilStreicher.pdf.

SUNDHOLM, G. "Proof-Theoretical Semantics and Fregean Identity Criteria for Propositions". *The Monist*, v.77, n.3, pp.294-314, 1994. Disponível em: https://doi.org/10.5840/monist199477315.

UFP (The Univalent Foundations Program). *Homotopy type theory: Univalent foundations of mathematics*, 2013. Disponível em: https://homotopytypetheory.org/book/.

ON THE IMPREDICATIVITY AND CIRCULARITY OF FREGE'S ANCESTRAL

João Vitor Schmidt[1]

ABSTRACT

Among Frege's major contributions to the philosophy of mathematics lies his famous definition for the Ancestral relation. Presented first in the 1879 *Begriffsschrift*, it has an important role in Frege's logicism, given that it provides the conditions for transforming the predecessor relation into a linear-ordered series. One of the objections for Frege's definition comes from its impredicativity and the circularity thus yielded. Such objection was pointed by Benno Kerry in 1887, and more recently by Ignacio Angelelli in 2012. In this article, I argue from a Fregean perspective that Frege's Ancestral is not circular, although its inevitably impredicative character.

Keywords: Gottlob Frege. Ancestral Relation. Impredicativity. Logicism.

RESUMO

Entre as principais contribuições de Frege à filosofia da matemática encontra-se a sua famosa definição de Ancestral de uma Relação. Apresentada primeiro na *Begriffsschrift* de 1879, tal definição possui um papel crucial no logicismo fregeano ao estabelecer as condições para transformar a relação predecessora em uma série linearmente ordenada. Uma das objeções à definição de Frege provém de sua impredicatividade e da circularidade resultante. Tal objeção foi primeiramente apresentada por Benno Kerry em 1887, e mais recentemente por Ignacio Angelelli em 2012. Neste artigo, argumento, a partir da perspectiva fregeana, que o Ancestral não é circular, não obstante sua inevitável impredicatividade.

Palavras Chave: Gottlob Frege. Relação Ancestral. Impredicatividade. Logicismo.

1. Logicism and Frege's Ancestral

Frege's definition of the Ancestral, as stated in the third section of the *Begriffsschrift*[2], is a second-order definition for the transitive closure of any first-order relation. For any first-order relation R and objects a and b, we say that Frege's Ancestral, denoted by $R^*(a,b)$, holds if, and only if, b has all the

[1] Mestre e Doutorando em Filosofia pela Universidade Estadual de Campinas – UNICAMP. Bolsista CNPq.
Email: joaovitorschmidt@gmail.com. ORCID: 0000-0003-3572-6306.

[2] Following the English translation in (FREGE, 1967). I reserve the German *Begriffsschrift* as denoting the 1879 book, and *concept-script* as denoting the logical notation presented in it.

hereditary properties that a has and that all immediate R-successors of a also have. Formally, this is

$$R^*(a,b) =_{\text{def}} \forall F\left[Her(F,R) \wedge \forall z(R(a,z) \rightarrow F(z)) \rightarrow F(b)\right]$$

with $Her(F,R)$ meaning "F is hereditary in R", formally

$$Her(F,R) =_{\text{def}} \forall x \forall y(F(x) \wedge R(x,y) \rightarrow F(y))$$

That is, F is a property that is passed along the objects satisfying the relation R. The name "Ancestral" has its origin in the fact that, if we consider R as denoting the parent relation, then R^* is the ordinary ancestral relation. It is also considered that Frege's definition is a logical reduction of this ordinary notion.

This was an important step for Frege's logicism since it provides the conditions to transform the predecessor relation into a linear-ordered series, being both transitive and trichotomous. The predecessor relation P is defined in Frege's *Grundlagen der Arithmetik,* §76[3] as:

$$P(m,n) =_{\text{def}} \exists F \exists x\left[F(x) \wedge n = \#F \wedge m = \#\left[z : F(z) \wedge z \neq x\right]\right]$$

Where we read that m precedes n if, and only if, n is the number of the objects satisfying a given concept and m is the number of such objects, except for one of them. Here, $\#$ is a number operator, denoting the number that is associated with the concept F[4]. With such relation, P^* denotes the series of objects that are correlated through P, being P^* both transitive and trichotomous. First, it's provable that

$$P^*(x,y) \wedge P^*(y,z) \rightarrow P^*(x,z)$$

Which is an instance of Frege's theorem 98 of the *Begriffsschrift* about the transitivity of the Ancestral, that provides the basic connection between

[3] Following Austin's translation in (FREGE, 1953). Henceforth quoted simply as *Grundlagen*.

[4] Frege introduces it with the abstract principle known as Hume's Principle in *Grundlagen*: $\#F = \#G \equiv F \approx G$, where $\approx$ is a bijection between the F's and the G's.

each and every element linked through R. Frege also offers the following two definitions: The Weak-Ancestral of R

$$R^{+}(x, y) =_{\text{def}} R^{*}(x, y) \vee x = y,$$

and the Single-Valuedness of a relation

$$Fun(R) =_{\text{def}} \forall x \forall y \forall z (R(x, y) \wedge R(x, z) \rightarrow y = z).$$

Using the former, Frege defines the natural numbers in terms of following the number zero in the series generated by the predecessor relation or being equal to zero, that is,

$$N(x) =_{\text{def}} P^{+}(0, x)$$

Following the fact that the predecessor relation is functional, *i.e.* single-valued[5], it can be proved as an instance of Theorem 133 of the *Begriffsschrift*:

$$N(x) \wedge N(y) \rightarrow \left(P^{*}(x, y) \vee P^{*}(y, x) \vee x = y\right),$$

which means that, for any pair of natural numbers, trichotomy holds for them. This yields that the series of natural numbers is linear-ordered.

The Ancestral also makes it possible to derive a general principle of induction, Theorem 81 of the *Begriffsschrift*,

$$\left[F(x) \wedge Her(F, R) \wedge R^{*}(x, y)\right] \rightarrow F(y)$$

and the principle of mathematical induction in the form[6]:

$$F(0) \wedge \forall v \forall w (N(v) \wedge F(v) \wedge P(v, w) \rightarrow F(w)) \rightarrow \forall z (N(z) \rightarrow F(z))$$

As a major result from all these facts, the Peano-Dedekind axioms for arithmetic are derivable in Frege's logic, assuming, of course, that the number-operator # is well-defined and Hume's Principle as a legitimate substitute

[5] This is declared in the *Grundlagen*, §72, proposition 5.

[6] Assuming the necessary corrections pointed out by Boolos and Heck in (BOOLOS & HECK, 2011)

for Frege's Basic Law V. Frege's definition of the Ancestral, then, has a substantial role in Frege's logicism.

Philosophically, Frege's Ancestral was his first argument against the necessity of kantian intuitions for ordered series and the foundations of Arithmetic, as well for the informativity of (second-order) logic. Kant's philosophical position towards numbers and arithmetic, as exposed in the first *Critique* (KANT, 1998), takes logic as a formal discipline, a *canon* of reason which does not have any content of his own. For this reason, Kant believed that pure intuitions were necessary to preserve both the content and the *a priori* character of the arithmetical science. In this scenario, Kant's argument for the well-ordering of arithmetic is the following: Numbers are to be regarded as symbolic constructions in time, one that depends upon the counting or adding homogeneous units in each interval. Since time, being uni-dimensional and unlimited, is a linear order by default, each number constructed preserves its position in the timeline. The well-ordering of the natural numbers is then an easy corollary for Kant.

Frege's position against Kant starts by defining a way to characterize linear orders without the pure intuition of time. This involves a criticism of the contentless character of Kant's notion of formal logic and the triviality of analytic judgments. Frege takes his concept-script logic to be contentful, and the formal results regarding the Ancestral about linear-ordered series to be his first step towards this goal. He explicitly mentions this in the *Grundlagen*, in regarding the proof of theorem 133 as being not only contentful but also analytic, remarking how "From this proof, it can be seen that propositions which extend our knowledge can have analytic judgements for their content" (FREGE, 1953, §91). This is evidence that already in 1879 Frege was not only endorsing logicism but also criticizing Kant's position, something that was only made clearer in the *Grundlagen*, five years later. All this through the Ancestral definition.

But Frege's logicism, despite all its merits, did not survive Russell's Paradox, following the inconsistency of Basic Law V of the *Grundgesetze der*

Arithmetik system[7]. The failure of Frege's axiom V was not just an isolated problem, but something that undermined mostly every attempt into the foundations of mathematics that used an unrestricted notion of sets or classes. Russell, too, was affected by it and attempted his own amendment in the years following the discovery of the paradox. The root of the problem, something that unrestricted comprehension axioms presuppose, was that any formula could define a set or class[8]. But Russell's Paradox shows otherwise. It cannot be the case that expressions like "$x \notin x$" define a set or class. Russell (1907) called such cases as *non-predicatives*, later labeled as *impredicatives*. They are distinguished from those expressions that can define a set as *predicatives*. But what exactly prevents such impredicative expressions from defining sets was better argued later by Poincaré[9]: the presence of a vicious circle, or the attempt to define an object in terms of a domain that already contains it. Russell's Paradox is a prime example where such a vicious circle leads to paradoxical situations, and for that reason, Russell's solution was to ban such unrestricted quantification, which evolved into his Ramified Type Theory. This was done based on the *vicious circle principle*: that no object (or set) should be defined in terms of itself, or the aggregate in which it is an element.

But are all impredicative definitions harmful? The question whether impredicative definitions should be accepted was intensively debated in the first half of the twentieth century, to the point of Carnap calling it the "The *most difficult problem* confronting contemporary studies in the foundations of mathematics" (CARNAP, 1964, p.49), especially because predicative principles, such as Russell's vicious circle principle and the ramified theory of types, were successful in avoiding paradoxes, but were inadequate for dealing with some portions of mathematics, more crucially with the real numbers[10]. But in

[7] Following Ebert and Rossberg's translation in (FREGE, 2013).

[8] Frege did accept comprehension principles for concepts in 1879, and functions in 1893. However, he did it not in terms of axiom schemes, but through an undeclared rule of uniform substitution. Nonetheless, it is only with the presence of Axiom V that comprehension is problematic. In axiomatic set-theory such as ZFC, in order to avoid such problems, a limited version of comprehension is adopted as an axiom schema of separation.

[9] See, for example, chapter IV in (POINCARÉ, 1963). For a discussion on this topic, see also (FEFERMAN, 2005).

[10] Carnap's quote continued: "How can we develop logic if, on the one hand, we are to avoid the danger of the meaninglessness of impredicative definitions and, on the other hand, are to

the philosophical side of the debate, a common wisdom was formed: that impredicativity was generally acceptable under a Platonistic background[11].

Regardless if such common wisdom is faithful to the debate, we can extend it back to Frege's Ancestral. Already in 1887, Benno Kerry pointed out that Frege's definition was impredicative, given its circular nature. Since Frege's strategy was to define natural numbers in terms of the Ancestral, this definition was doomed to fail, according to Kerry. More recently, Ignacio Angelelli (2012) argued the same, adding that such circularity also undermines Frege's reduction of the very notion of a series in logical terminology. My main goal here is to defend that, from Frege's point of view, *i.e.,* his realist background, the definition of the Ancestral is not as harmful as considered by Kerry and Angelelli. As a bonus, and since Frege never answered Kerry, I'll argue that Frege could have been the first realist defense of impredicativity[12].

Towards this goal, I'll restate both Kerry's and Angelelli's arguments against Frege's definition to then recapitulate Frege's reduction and some of his philosophical justification for it. This will set the course for the second point: that the circularity is true only under a questionable assumption about quantification in Frege's concept-script and logic in general. It will help us understand why such circularity is not a problem for Frege, and perhaps why he never answered Kerry's criticisms. After comparing other famous answers to Russell's vicious circle principle with the Fregean one here proposed, I'll conclude that, at least in this scenario, Frege's Ancestral is not circular, although still impredicative.

---

reconstruct satisfactorily the theory of real numbers?" (CARNAP, 1964, p.49). For instance, quantification over the reals was seriously limited by Russell's ramified theory, forcing some reinterpretation of basic expressions such as "for all real numbers." Moreover, classical mathematics is seriously limited if not by assuming some impredicative instances.

[11] This is summarized and evaluated by Parsons (2014). Some of the authors discussed will be assessed later.

[12] This, I should mention, must be read regarding only the impredicativity of the Ancestral. It is well known that Frege's logic is impredicative also in the sense of adopting full comprehension schemes for formulas, which, adding axiom V, is responsible for the inconsistency of Frege's second-order logic. But no inconsistency is known regarding the impredicativity of the Ancestral.

2. The impredicativity of Frege's Ancestral

Textually, Frege's definition of the Ancestral states that x is the R-ancestor of y (*i.e.*, y follows x in the R-series), just in case y has all F-hereditary properties shared by all R-successors of x (*i.e.*, all objects that follows x in the R-sequence). This is a second-order definition since it quantifies over the domain of concepts, which in Frege's case, is unrestricted. In addition to the more famous debate on his theory of concepts, Benno Kerry also criticized Frege over this definition. This was done essentially in 1887 in *Über Anschauung und ihre psychische Verarbeitung* (KERRY, 1887, p.295), where he states that:

> Now, this criterion is to begin with of dubious value because there is not a catalogue of such properties, hence one is never sure that one has examined the totality of them. Moreover, there is the crucial fact that, as the author himself has proved [in a footnote Kerry cites *Begriffsschrift,* Theorem 97], of the properties that are hereditary in the f-series is also the following: to follow x in the f-series. Thus, the determination of whether y follows x in the f-series, according to the definition given for this concept, depends on whether, in addition to a lot of other things on hereditary properties in general, one knows, in particular, about the hereditary property "being a descendant of x", that y has it or not. It is clear that this circle should totally prevent from saying, in Frege's sense, that any y follows x in an f-series[13].

In justifying if $R^*(x, y)$ holds or not, it is required that, for every hereditary property F, one can decide whether F is one of the properties that is passed along from x to y or not. Kerry's point is that since Frege's definition quantifies over all hereditary properties, there could be one for which this task is uncertain. The case in question, as Kerry quotes, is Frege's theorem 97:

$$\forall u \forall v\left(R^*(x,u) \wedge R(u, v) \rightarrow R^*(x, v)\right)^{14}$$

This theorem states that the property $\left[z: R^*(x, z)\right]$, which we can read as "following x in the R-series", is hereditary in R. This is the same as saying that the property "being x descendant" is something that every descendant of x pass along the parent relation. For example, if Gottlob is Karl's descendant,

[13] The translation is Angelelli's (ANGELELLI, 2012). Where he translates "being a descendant of x", Kerry actually writes simply "to follow x", since he, as Frege, does not name such relation as Ancestral.

[14] I avoided here the *definiendum* of the Hereditary Property, showing only the *definiens*.

then Alfred, Gottlob's son, is Karl's descendant as well. The problem pointed out by Kerry is that to determine whether Alfred is Karl's descendant, we have to find all such hereditary properties and check if Alfred has them. But in this process, the above property would require to determine if Alfred is Karl's descendant. Hence the circularity.

This circularity was later reassessed by Ignacio Angelelli in *Frege's Ancestral and his Circularities* (ANGELELLI, 2012). He presented two versions, the first one being Kerry's, as stated above. The second one aims to show how Frege's intended reduction of the notion of a series failed as well. Simply put, a series is any connection between a set of elements that is at least transitive. It states that there is a "chain" between one object to another given a finite number of steps. Frege did not name such notion as the Ancestral, but such relation is equivalent: both are transitive connections, or paths, between one element (the ancestor) to another (the descendant). This is what Angelelli defines as the Ordinary Ancestral, henceforth labeled as **(OA)**.

Frege's definition was a reduction of **(OA)** into logical terminology. This is what he argues in (FREGE, 1967, p.5): "My initial step was to attempt to *reduce* the concept of ordering in a sequence to that of *logical* consequence, so as to proceed from there to the concept of number"[15]. This is also, according to Angelelli's interpretation, what Frege maintains in defining the Ancestral, in (FREGE, 1953, §79), by saying that the *definiens* "is to mean the same as" [*sei gleichbedeutend*] the *definiendum*. For that reason, Angelelli concludes that:

> It seems natural do interpret the product of the reduction as intended to replace the initial notion. In alternative terms, it seems natural to construe the Fregean Ancestral as an *analysans* that replaces, in Frege's project, the *analysandum* (the common ancestral). Such would be the analysis interpretation. (ANGELELLI, 2012, p.478)

Since Frege was also interested in developing a logical system capable of replacing the natural language, his logical definition of the Ancestral, henceforth **(FA)**, should be taken as a replacement of **(OA)**. But was Frege's intended reduction successful? According to Angelelli, no.

[15] The first emphasis is mine. The second is Frege's.

Kerry and Angelelli's versions of the circularity have the same basic principle: that **(FA)** is circular given its unrestricted quantification over properties. I focus here on Angelelli's versions, with the following fictional situation added by him: a certain Fritz is trying to convince a jury that he is Karl's descendant to inherit Karl's money. Since Fritz doesn't have the necessary documents, he is tempted to quote Frege's definition, *viz.,* that he is Karl's descendant if, and only if, he has all hereditary properties shared by Karl's descendants. The jury then asks him to check whether Frege's definition could help him with two specific properties: "being the Fregean-descendant of Karl" and "being the ordinary-descendant of Karl." Recall that **(FA)** is $R^*(x, y)$, and consider now $R^o(x, y)$ as denoting that x is the ordinary ancestor **(OA)** of y[16]. The properties questioned by the jury are $[z: R^*(k, z)]$ and $[z: R^o(k, z)]$ respectively, where k is a name for Karl. Assuming that Fritz wants to check such properties, the first one generates Kerry's circularity. The second exemplifies Angelelli's argument against Frege's reduction of the Ancestral, what he calls the analysis interpretation. More precisely, and assuming that k and f are names for Karl and Fritz, the first is the following:

Kerry's Circularity:

(a) In order to prove that $R^*(k, f)$ holds, one has to show all F properties which are hereditary in R such that, if $\forall z(R(k, z) \rightarrow F(z))$ then $F(f)$;

(b) $[z: R^*(k, z)]$ is such property;

(c) Then, one has to show that f has the property $[z: R^*(k, z)]$, *i.e.*, that $R^*(k, f)$ holds;

(d) This is circular, hence, $R^*(k, f)$ cannot hold.

The second circularity is an argument against the Analysis Interpretation. It goes as follows:

[16] This simply means that $R^o(x, y)$ should be read intuitively as the existence of a path from x to y.

Angelelli's Circularity

(a) Assume that **(FA)** is a reduction of **(OA)**;

(b) From (a), to show that $R^o(k, f)$ holds, one has to show that $R^*(k, f)$ holds;

(c) From **(FA)**, to prove that $R^*(k, f)$, one has to show all F properties which are hereditary in R such that, if $\forall z(R(k, z) \rightarrow F(z))$, then $F(f)$;

(d) $[z: R^o(k, z)]$ is such property;

(e) Hence, one has to prove that f has the property $[z: R^o(k, z)]$, that is, prove that $R^o(k, f)$ holds;

(f) This is circular. Therefore, **(FA)** is not a reduction of **(OA)**.

Both circularities can be generalized for any x and y. The first was already discussed above. The second one goes between (b) and (e). It simply states that if one wants to prove that someone is his ordinary ancestor in terms of Frege's definition, at some point, he would have to prove that he is the ordinary ancestor of that person. Hence, the proof would require the conclusion as one of its premises. The consequence of such circularity is that to **(FA)** successfully reduce **(OA)**, **(OA)** should be stated in terms of **(FA)** alone. Still, no such reduction is possible since every time we want to show that **(OA)** holds in terms of **(FA)**, we are obligated to prove **(OA)** in advance.

Angelelli's conclusion, then, is that at best **(FA)** is an enrichment and a generalization of **(OA)**, not an analysis or reduction. In his words:

> The enrichment occurs through the "discovery" of the property of being hereditary that many properties have [and] includes the focusing on, and helps towards the demonstration of the formal properties of the ordinary ancestral, *e.g.*, transitivity, which as Frege points out is what leads to the logical understanding of arithmetical induction. The generalization is accomplished in that the ancestral's underlying relation as such is conceived in a most abstract fashion. (ANGELELLI, 2012, pp.498-499)

Since **(FA)** quantifies unrestrictedly over the domain of properties, the ordinary ancestral still appears in its scope. For that reason, **(FA)** cannot accurately substitute **(OA)**. The alleged enrichment of **(FA)** has already been dis-

cussed above. It is not only transitive, but given the suitable conditions **(FA)** is also trichotomous and fundamental for the natural number series and the principle of mathematical induction. The generalization is due solely to the fact that the Ancestral is actually not a first-order relation *per se*, but a second-order property that can be applied to any first-order relation. Either way, Frege's definition is impredicative, as the first circularity clearly shows.

But yields the impredicative nature of **(FA)** a circular definition? Is it as harmful as Kerry and Angelelli supposed? At least in Fregean standards, it isn't. Both arguments have, at least, one problematic premise. Added with Frege's philosophical motivations, this might help to understand and avoid both circularities.

3. Frege's Reduction: Logicism and Realism

As mentioned earlier, logicism was Frege's primary motivation from the beginning, one that could very well be stated as the attempt of freeing arithmetic from intuitions. This was done mainly in two fronts: first, in showing that arithmetical concepts are reducible into logical terminology and second, in showing that arithmetical modes of inference are reducible to logical modes of inference. Angelelli is well aware of these motivations. He recalls two crucial points: (1) that for Frege any consideration about features of particular cases of **(FA)** are not essential and (2) that the proof of an instance of $R^*(x, y)$ should not be an enumeration of each point of the chain that starts from x and leads to y. These are essential points that Angelelli rightfully mention, but failed to link to the circularity problem. The first thing, argued in the *Begriffsschrift*, is the generality of **(FA)**:

> The propositions about sequences developed in what follows far surpass in generality all similar propositions which can be derived from any intuition of sequences. Therefore, if one wishes to consider it more appropriate to take as a basis an intuitive idea of sequences, then he must not forget that the propositions so obtained, which might have somewhat the same wording as the ones given here, would not state nearly so much as these because they would have validity only in the domain of the particular intuition upon which they were founded. (FREGE, 1967, §23)

It follows that Frege's definition is one that generalizes the basic notion of a (transitive) series. Thus, examples as the ordinary ancestral relation holding between human beings, the relation between a number following another in the natural number series, or any other partially ordered series are instances of Frege's definition. Particular cases of **(FA)** can be about objects that are only apprehended from some intuition or none at all, but **(FA)** does not have such restrictions. This is confirmed by the fact that theorems regarding it are proved formally, or as Frege would put it, from "pure thought" alone (FREGE, 1967, §23). The generality of the definition is again mentioned in (FREGE, 1953, §80): "since the relation R has been left indefinite, the series is not necessarily to be conceived in the form of a spatial and temporal arrangement, although these cases are not excluded." Following this passage, Frege starts arguing against what we might consider a step-by-step proof procedure for the Ancestral. This is: "if starting from x we transfer our attention continually from one object to another to which it stands in the relation R, and if by this procedure we can finally reach y, then we say that y follows in the R-series after x" (*ibidem*)[17]. From this procedure, if Fritz wants to prove that he is Karl's descendant, he just has to prove that he is the son of a son of ... a son of Karl, something like:

$$\exists x_1, \ldots, \exists x_n \left(P\left(k, x_1\right) \wedge, \ldots, \wedge\, P\left(x_n, f\right) \right)$$

where P is the usual parent relation. A step-by-step proof would require that each link of the chain between Karl and Fritz be proved to exist. Frege does not rely on this kind of proof. He continues by saying that "this describes a way of *discovering* that y follows, it does not define what is *meant* by y's following" (*ibidem*)[18]. Surely, a discovery would be enough to *show* that Fritz follows Karl in the Ancestral relation, but is this enough for a proper logical analysis? Fritz ancestry does not depend upon his proof in front of the jury, it depends on some conditions that obtain by virtue of him *being Karl's descendant*, which, in Frege's analysis, depends on him having some required pro-

[17] Frege uses an uppercase Φ to express the relation that here, and henceforth, is denoted by the usual roman uppercase R.

[18] The emphasis is mine.

perties that are a consequence of the fact that such relation between Karl and Fritz obtains. Thus, for a proper logical analysis, we should not take the discovery as relevant, neither should we define a concept or relation that depends on it. As Frege argues:

> Whether y follows in the R-series after x has in general absolutely nothing to do with our attention and the circumstances in which we transfer it; on the contrary, it is a question of fact, just as much as it is a fact that a green leaf reflects light rays of certain wave-lengths whether or not these fall into my eye and give rise to a sensation, and a fact that a grain of salt is soluble in water whether or not I drop it into water and observe the result, and a further fact that it remains still soluble even when it is utterly impossible for me to make any experiment with it. (*ibidem*).

This relates to Frege's realism: that there are facts independent of our way of apprehending them. This is the case for arithmetic, since numbers are, as he regards them, self-subsisting objects. Likewise, arithmetical propositions are true independently of our way of regarding them.

This realist approach towards logic and arithmetic also assumes that properties (concepts) are objective and independent entities. They are things that we *grasp*, not things that we construct:

> [...] a concept is something objective: we do not form it, nor does it form itself in us, but we seek to grasp it, and in the end we hope to have grasped it, though we may mistakenly have been looking for something where there was nothing. (FREGE, 1984, P.133).

The same goes to definitions, as they do not have the power to create anything, only to grasp something:

> Just as the geographer does not create a sea when he draws borderlines and say: the part of the water surface bordered by these lines I will call Yellow Sea, so too the mathematician cannot properly create anything by his definitions (FREGE, 2013, p.XIII)

Fregean definitions are non-creative: they are not responsible for putting something into existence but instead are useful conventions used for inferential purposes. A Fregean definition *grasps* a concept; it does not create it. Thus, in Frege's realism, there is no problem in defining something in terms of the totality in which is itself a member, given that the object defined - in the case of the Ancestral, the second-order concept in which binary relations fall – exists independently of the definition.

> on the one hand, we can ask by what path a proposition has been gradually established; or, on the other hand, in what way it is finally most firmly establishable. Perhaps the former question must be answered differently for different people. The latter [question] is more definite, and its answer is connected with the inner nature of the proposition under consideration. (FREGE, 1967, p.103)

The realist stance also relates to Frege's anti-psychologism: roughly that we cannot take the mental representation of an object as the object itself. At least as far as arithmetical propositions are concerned, the genesis of our mental representations about some fact has nothing to do with the justification for its truth. This is argued in the very beginning of the *Begriffsschrift*:

> He then concludes that it's "[...] not the psychological mode of origin, but the most perfect method of proof underlies the classification" (*ibidem*). The point was also made in the introduction of *Grundlagen* as a *dictum*:
>
> > Never let us take a description of the origin of an idea for a definition, or an account of the mental and physical conditions on which we become conscious of a proposition for a proof of it. [...] a proposition no more ceases to be true when I cease to think of it than the sun ceases to exist when I shut my eyes. (FREGE, 1953, p.xviii)

All this reminds us of how Frege distinguished between the discovery that a proposition is true from the justification for it being true.

Returning to Fritz's case, one should expect to prove his heritage by showing each step of the chain between him and Karl. But Frege's realism towards truth assumes that this fact, whether Fritz is or isn't Karl's descendant, does not change according to his defense for it. He is Karl's descendant or not, regardless if the jury is convinced with his speech. Frege himself is saying that:

> What I have provided is a criterion which decides in every case the question Does it follow after?, wherever it can be put; and however much in particular cases we may be prevented by extraneous difficulties from actually reaching a decision, that is irrelevant to the fact itself. (FREGE, 1953, §80)

According to Frege's philosophical recommendations, $R^*(x, y)$ cannot rely upon a step-by-step proof, the type of which shows each link of the connection between x and y. Theorem 98, about the transitivity of the Ancestral,

could be equally proved in this intuitive way. If there is a path between x and y, and another from y to z, we could then prove that there is another from x to z, by simply starting from x, reaching y and moving along until we finally reach z. But should the proof depend on such a procedure? Certainly not:

> We have no need always to run through all the members of a series intervening between the first member and some given object, in order to ascertain that the latter does follow after the former. Given, for example, that in the R-series b follows after a and c after b, then we can deduce from our definition that c follows after a, without even knowing the intervening members of the series. (*ibidem*).

As we saw, Frege's intended to offer a definition that makes possible to prove important facts about the Ancestral in a purely logical way, with maximal generality and without any need for intuitions in its proofs. This is **(FA)**. **(OA)**, on the other hand, is the intuitive notion that Frege wants to avoid, one that requires a step-by-step proof relying on particular cases of application. Thus, and Angelelli is right about this, **(FA)** is a reduction of **(OA)**, a generalization and certainly an enrichment, one that seeks to provide for the common notion a formal and precise formulation that is purely logical. But Angelelli's and Kerry's argument for the circularity made use of an assumption that is very much akin to what Frege is trying to rule out with **(FA)**.

4. The circularity revisited

Why is Fritz incapable of proving his desired heritage of Karl's by simply quoting **(FA)**? Because **(FA)** doesn't work that way. By asking Fritz to prove whether he has the property "being Karl's descendant", the jury is asking for the exact step-by-step proof that Frege wanted to shun. But instead of asking proof for each link of the ancestor to the descendant, they ask for each hereditary property, those that are passed along from one to the other, to be proved in advance.

This is what happens with both arguments for the circularity of **(FA)**. The assumption is that to verify the truth of a quantified sentence like $\forall xF(x)$, one has to show first that F holds for every object in the domain. Thus, a universally quantified sentence would require a justification for each of its ins-

tances to be true. Even more, it would require that the desired sentence should be accepted as true *only* if, in principle, we verify each instance in advance. But, of course, this would get us into trouble if the quantified domain is not finite, something that is not clear in the case of **(FA)**. In this scenario, sentences like "every natural number is either even or odd" would require more than is humanly possible to verify to be true. This is not what quantification is about. However, this is still presumed if such circularity is to be derived from **(FA)**: only in this scenario, the question whether b has the property $\left[x : R^{*}(a, x)\right]$ can be raised *prior* the justification for $R^{*}(a, b)$.

This is what at least one premise of each argument presupposes to be valid: premises (a) and (c) from the first and second arguments, respectively. They assert that to prove $R^{*}(a, b)$, one has to show all F properties which are hereditary in R such that, if $\forall z(R(a, z) \rightarrow F(z))$, then $F(b)$. This is precisely the problem that Frege had answered in the *Grundlagen*. One needs no verification of each step of the sequence to prove that $R^{*}(a, b)$. Analogously, one needs no checking of each hereditary property to prove that $R^{*}(a, b)$. In this case, both Kerry and Angelelli seem to be taking a verification for logical justification.

First, we have a step-by-step proof of the Ancestral relation that Frege ruled out, *i.e.* the checking of each link of the series in order to verify the connection between the ancestor and the descendant. This is an epistemic verification, *viz.* a discovery. In the same way, Angelelli and Kerry seem to be assuming an epistemic verification for the application of **(FA)**: the checking of each hereditary property that the ancestor passes through R and the subsequent checking of the same properties in the descendant. It's safe to say that Frege would rule that out too.

There seems to be a confusion in both circularity arguments regarding the role of the universal quantifier and its logical justification. Frege never answered Kerry about the circularity problem, but Russell had in the appendix dedicated to Frege's philosophy in his *Principles of Mathematics.* About Kerry's argument, Russell states:

> This argument, to my mind, radically misconceives the nature of deduction. In deduction a proposition is proved to hold concerning *every* member of a class, and may then be asserted of a particular member: but the proposition concerning *every* does not necessarily result from enumeration of the entries in a catalogue. (RUSSELL, 1996, p.522)

The same can be applied to Angelelli's argument. He even mention Russell's passage but doesn't discuss the response in his paper. He also mention Carnap's response, but also does not consider it. Carnap's point is in favor of Russell and Frege:

> [...] in order to demonstrate the truth of a universal sentence, it is not necessary to prove the sentences which result from it by the substitution of constants; rather, the truth of the universal sentence is established by a proof of that sentence itself. The demonstration of all individual cases is impossible from the start, because of their infinite number, and if such a test were necessary, all universal sentences and all indefinite predicates (not only the impredicative ones) would be irresoluble and therefore (by that argument) meaningless. (CARNAP, 1937, p.163)

Then, not just this proof procedure is untenable but is impossible in cases where the domain in question is infinite since such checking procedures are finite operations.

Even though Angelelli is aware and correct about Frege's intentions, I take that these points are sufficient to disqualify both circularities under Frege's background, or at least to raise significant issues on it. In Angelelli's version, in contrast to Kerry's, the focus is on the proof that $\left[x : R^{o}(k, x)\right]$ is a hereditary property in R (premise (d)). But as I see it, this can be answered in the same way. The question of whether Fritz satisfies or not property $\left[x : R^{o}(k, x)\right]$ does not depend on any checking. Either $R^{o}(k, f)$ is the case or it isn't regardless of Fritz's arguments. In the fictional story, it's pretty clear that Fritz goes to the court to show his heritage, not to determine the truth of it. Frege's Ancestral is insufficient for the task only because it is not a definition designed for such task in the first place. The proper question is this: if Fritz is Karl's descendant, does **(FA)** holds for them? I see no reason to think that it doesn't[19].

[19] The other way to put it is by asking whether Frege's ancestral is extensionally correct in capturing the cases where there exists a finite path between the ancestor and the descendant. Heck (2016) proves that the answer is yes, but this requires some sort of mathematical in-

Fulugonio (2008) also offers an argument in favor of the circularity reading. Her arguments are very similar to Angelelli's[20] as it reiterates that the circularity undermines Frege's project completely. She argues that:

> The only supposition in Kerry's criticisms to the Fregean definition of succession is that Frege's construction, and the definition of succession, in particular, has a defined gnosiological motivation, one that Frege certainly explicit in many opportunities throughout his work. From that, if, as part of his project, Frege intends to clarify the notion of succession, his elucidation is - at least - unsatisfactory if it demands, among other things, that it is known what it attempts to be clarifying. (FULUGONIO, 2008, p.9)[21]

Fulugonio's point is precisely Angelelli's second circularity argument. Frege indeed had a gnosiological program in mind: to show that all arithmetical truths are analytic, not to mention *a priori*, which in turn should prove that all our knowledge on numbers is logically justified. This program includes the so-called reduction of ordinary notions to logical definitions, as the Ancestral. But the crux of Fulugonio's point is the same as before: the hidden assumption that any universally quantified proposition demands a verificational proof, *i.e.,* a proof for each instance. But our conditions for verifying any number of cases – finite or infinite – should not be taken as responsible for the proof of a quantified proposition. Frege did make this point on multiple occasions. He warned in the *Grundlagen* and earlier in the *Begriffsschrift* not to take the *being* true as *taken* to be true. The analyticity of arithmetic should be a consequence of its logical grounds, not the other way around. As far as the Ancestral is concerned, definitions and proofs should not be taken as verificationally motivated.

---

duction on the number of steps that go from each node of the parent relation. This is worrying because Frege's ancestral supposed to derive a principle of mathematical induction, not to depend on it for its extensional correctness. But Heck also shows that one can define the Ancestral in a way that avoids such circularity and that is still equivalent to (FA).

[20] And in her defense, came earlier than Angelelli's paper.

[21] In the original: "El único supuesto presente en la crítica de Kerry a la definición fregeana de sucesión es que toda la construcción fregeana, y la definición de sucesión en particular, tiene una aspiración gnoseológica definida, aspiración que por cierto Frege explicita en reiteradas oportunidades a lo largo de toda su obra. De modo que, si como parte de su proyecto lo que pretende Frege es elucidar la noción de sucesión, su elucidación es –cuanto menos– insatisfactoria si ella exige, entre otras cosas, que sea ya conocido aquello que se quiere elucidar". The translation is mine.

Furthermore, Fulugonio argues that Russell's defense of Frege's definition of the Ancestral missed one important point in not discussing the problematic property. As she says, "[..] 'to follow x' is precisely one of the hereditary properties in which we find ourselves in a vicious circle." (FULUGONIO, 2008, p.8)[22]. But Russell's response goes to the heart of the problem: only by taking an incorrect account on quantification that such property becomes problematic as Kerry's supposed in the first place. In his short response to Angelelli, Heck (2016, p.101) added that "Russell is making an elementary logical *cum* epistemological point: a universal generalization does not have to be derived from the conjunction of its instances, and knowledge of a universal generalization need not rest upon knowledge of its instances".

It should also be noted that Fritz's case is a particular one: a specific application of **(FA)** to a specific individual. For that matter, it's hard to imagine that Fritz could prove $R^{*}(k, f)$ without a likely empirical (*viz.* intuitive) proof. And without some grounds, he cannot. It seems that **(FA)** is not suitable for helping his case, but again, this was not Frege's interest either. Heck also adds that "the power of Frege's definition shows itself not in particular cases but in results like [...] the *generalizations* that it allows to us to prove" (HECK, 2016, pp.101-102). Take, for example, the transitivity of the Ancestral (Theorem 98). Can we prove that the ordinary ancestral relation is transitive regarding every application in an intuitive way? Can we show that if $n<m$ and $m<p$, then $n<p$? If n, m, p are natural numbers then the number of nodes between n and p is finite. The task might be difficult, but is still possible. But the same cannot be said if n, m, p are extended to the Reals, since the number of nodes between n and p are not finite in this case.

Frege's intentions were with his logicist account of arithmetic, and in this particular case, no "checking" procedure is possible, if numbers are to be regarded as logical objects. Otherwise, propositions about numbers would be empirical ones. As Russell (1996, p.522) completes, "Kerry's argument, therefore, is answered by a correct theory of deduction; and the logical theory of arithmetic is vindicated against its critics". Russell also pointed out that this is

[22] In the original: "[...] que "seguir a x" es precisamente una de tales propiedades hereditarias, con lo cual nos encontramos ante un círculo vicioso". The translation is mine.

very akin to Mill's objection to Barbara inferences, but can be likewise answered. He concludes that a

> general proposition can often be established where no means exist of cataloging the terms of the class for which they hold; and even, as we have abundantly seen, general propositions fully stated hold of *all* terms, or, as in the above case, of *all* functions, of which no catalogue can be conceived. (RUSSELL, 1996, p.522)

For which Angelelli's argument can be likewise answered.

The catalog of all functions surely would be problematic for its cardinality, but for matters of both the circularity arguments, any infinite (countable) domain would already offer serious problems. Given that R^* is not restricted to any specific domain of interpretation, applications in domains already known to be infinite would be problematic. Considering the case of the natural numbers, it's clear that they satisfy an infinite number of hereditary properties. For instance, for a given pair of natural numbers (m, n), there exists a natural number p such that $m < p$ and $n < p$. Hence, 'being smaller than p' is undoubtedly a property shared by m and n in the predecessor relation. Since there are infinitely many p's, there would be infinitely many such properties.

This would not be a problem if, in principle, the domain of quantification were finite. But we are not even sure how the initial list of properties can be finite in the first place. Because if it is, we can still expand it indefinitely. Assume, for example, that there is such a finite catalog of properties and that **(FA)** is justifiably true just in case there is a way to verify each of the properties in it. This means that $R^*(a, b)$ holds if, and only if, for each $F_{i \leqslant n}$ that is hereditary, for a finite n, it is verified that if all R-descendants of a has F_i, then $F_i(b)$. This means that b inherits the hereditary properties of a. Assume also that all F_i's are decidable, that is, no circularity in Angelelli's terms is derivable from the list[23]. Thus, in theory, if every property is successfully verified, $R^*(a, b)$ is said to be verified as well. But why should we settle for the initial list? If b satisfies $F_i(x)$ then certainly b also satisfies $F_i(x) \vee \varphi$ for any new formula φ not originally in the list. Also, if b satisfies any two formulas in

[23] This implies that the property $[x : R^*(a, x)]$ is not in the list.

the list, then it certainly satisfies the conjunction, implication and the biconditional of those same formulas.

From this assumption, it is possible to expand the initial finite list endlessly, starting only from those formulas of properties known to our original list and adding every new formula obtainable from the closure of those formulas under the relevant logical operators. Generally, if $F_i(x)$ and $F_j(x)$ are in the list, for $i, j \leqslant n$ and n being the number of entries, then, $F_i(x) \vee \varphi$, $F_i(x) \wedge F_j(x)$, $F_i(x) \rightarrow F_j(x)$, $F_i(x) \equiv F_j(x)$ are all in the new list. Clearly, if b satisfies every property of the finite initial list, then b also satisfies all properties obtained in the new list, given that every new entry is also hereditary.

This procedure "generates" new properties. Even if we have only one property in the list, *i.e.,* if $n=1$, the new list will be expanded infinitely, forming every iteration of formulas from $F_1(x)$. The point is that we have no means to reject this infinite expansion if we assume that such a list exists to be verified. Even more, we have no good arguments to deny that this expanded list wasn't the one that we already had in the first place. Thus, if we are to assume a "checking" procedure, which is limited by our finite conditions, we would never check all of them. Once again, any quantification into an infinite domain would be untenable. The Fregean response would be that any proof of R^* should not be restricted to a finite checking whatsoever.

In Fritz's case, we could easily consider that even if he can show every decidable hereditary property that is passed from Karl's descendants to him, the jury could easily form a new list of hereditary properties from the previous one, thus extending Fritz task into infinity. But of course, this rather ordinary procedure is not the only way one can prove that $R^*(a, b)$, or Fritz to prove that he is Karl's descendant[24]. Fritz would easily prove that he has this property if he had grounds to show his heritages. If he is eager to use **(FA)**, these grounds must be logically justified. But it is hard to believe that this is what matters in Fritz's case. If the question was to decide whether for example $P^*(0,3)$ holds, surely there must be a logical way to check it. But Fritz is not a logical object, so there would be no pure logical form to decide whether he is

[24] The same point is made by Heck (2016).

Karl's descendant or not. Angelelli argues that **(FA)** is a reduction to the sense that it's supposed to substitute **(OA)**, to the extent that Frege's intended to provide a more precise language than the ordinary one. Frege's intention, as I see it, certainly was to provide better definitions for ordinary concepts, but *insofar* as they are necessary for mathematical purposes. In this sense, there is no reason to expect from Frege the intentions that Fritz has in the fictitious example that Angelelli provides.

Finally, if we expect Fritz to prove his heritages only in the presence of verification of each instance of the parent relation from him to Karl, Fritz would have problems to show, for example, that if he is the son of someone who is Karl descendant, he is Karl's descendant as well. But, providing these cases, it is clear that he would be, even without a demonstration of the complete path from his father to Karl. If Fritz had those two premises proved, he could prove that he is Karl's descendant with the aid of Theorem 96 of the *Begriffsschrift*:

$$R^*(x, y) \wedge R(y, z) \rightarrow R^*(x, z)$$

Likewise, should one prove that $\left[x : R^*(a, x)\right]$ is a hereditary property by looking to every member of the R-series? Certainly not. It can be easily proved from the definition alone. And Theorem 97 of the *Begriffsschrift* does exactly that.

5. Realism and Impredicativity

As we may conclude, there is no reason to embrace any version of Russell's vicious circle principle within the Fregean realist framework. In fact, the vicious circle principle did not receive full acceptance, as not everyone was in full agreement with the complete ban of such definitions, given that not every impredicative case is harmful or at least problematic. For Ramsey (1931), definitions are just a way of describing or putting into symbols the function or relation defined, since their totality already exists independently from the definition. Ramsey adopted the Wittgensteinian perspective of reading a quantified expression like $\forall F : F(x)$ as the logical sum of all propositi-

ons of form $F(x)$. Given, in his terms, "[…] our inability to write propositions of infinite length, which is logically a mere accident" (RAMSEY, 1931, p.41), there is no problem in describing them in terms of the totality in which they may be a member. In the now-famous example, he stated that

> [...] to express it like this [...] is merely to *describe it* in a certain way, by reference to a totality of which it may be itself a member, just as we can refer to a man as the tallest in a group, thus identifying him by means of a totality of which he is himself a member without there being any vicious circle (*ibidem.*)[25].

Thus, Ramsey agrees with the Fregean view that properties are not created, but grasped through definitions. A different stance was proposed by Carnap, since he did not accept the existence of such totality as given, even labeling Ramsey's position as "theological mathematics" (CARNAP, 1964, p. 50). As mentioned earlier, Carnap did recognize that universally quantified expressions need not be verifiable in terms of individual instances:

> If we reject the belief that it is necessary to run through individual cases and rather make it clear to ourselves that the complete verification of a statement about an arbitrary property means nothing more than its logical (more exactly, tautological) validity for an arbitrary property, we will come to the conclusion that impredicative definitions are logically admissible. (CARNAP, 1964, p.50).

He then continued by concluding that, even though it might be difficult to prove some instances of an impredicatively defined property, or even impossible, this difficulty or impossibility may not be a consequence of the impredicativity itself. Carnap's adopted a pragmatic perspective, as he elaborates more in *Logical Syntax of Language*. Accepting or not an impredicative definition is "[...] a question of choosing a form of language — that is, of the establishment of rules of syntax and of the investigation of the consequences of these" (CARNAP, 1937, p.164). In other words, it is just a matter of tracking the consequences and avoiding the problematic ones.

Perhaps more directly in tone with the Fregean solution here proposed was Gödel. He famously argued that impredicative definitions are problematic only by assuming what he calls a "constructivist" approach towards definitions:

[25] The emphasis is mine.

> [...] it seems that the vicious circle principle [...] applies only if the entities involved are constructed by ourselves. [...] If, however, it is a question of objects that exist independently of our constructions, there is nothing in the least absurd in the existence of totalities containing members which can be described (*i.e.*, uniquely characterized) only by reference to this totality. (GÖDEL, 1944, p.128)

So Gödel took a very similar standpoint to that of Frege, by assuming the objective existence of properties, classes, and the like. He went even further by taking their existence on a par with physical objects: "Classes and concepts may, however, also be conceived as real objects […] the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence" (GÖDEL, 1944, p.128).

Ramsey, Carnap, and Gödel are good examples of the opposite view from that of Russell and Poincaré's about the vicious circle principle. But Ramsey and Gödel are more representatives of the realist answer that Frege could have given to Kerry's objections, if he had provided one. Parsons (2014) discuss the positions that Ramsey, Carnap, and Gödel and other key authors such as Hilbert and Bernays. His conclusion goes against the conventional wisdom that the historical defense of impredicative definitions always goes along realist lines. Carnap certainly was one that denied some of the realist presuppositions and still accepted impredicative definitions. Moreover, the acceptance of classical mathematics inevitably forces one to accept impredicativity, and this is not exclusive of realist philosophers. Thus, the acceptance of impredicativity does not imply the adoption of realism. But the other direction might be the case: accepting a realist ontology, at least from the historical point of view, seems to imply the acceptance of impredicativity definitions. At least I know of no historical counter-example.

But in Frege's case, the implication is surely true, as I've tried to argue. The realist stance is an answer to the Fritz case. First, it is clear that Fritz is trying to show his heritage, not to create them, given that he is Karl's descendant regardless of his argumentation. $R^{*}(k,f)$ holds or not, even if Fritz cannot show that he has all hereditary properties passed from Karl to his descendants. Second, there is an infinite number of such properties, meaning that

Fritz would never be able to show all of them in the first place, assuming such constructivist reading about the Ancestral. The same line of thought follows for any application of the Ancestral, whether about Fritz, Numbers, or any set of objects describing a linear-ordered series.

6. Conclusion

The story is that Frege's definition is definitely impredicative, given that $R^*(a, x)$ is in the scope of the second-order quantifier in R^*, and Kerry must be praised as perhaps the first one to realize the problem of impredicativity in logic. The problem is also related to the Fregean definition of Natural Number. Recall that n is a natural number just in case $P^+(0, n)$ holds for n. But this is also a hereditary property: a successor of a natural number is itself a natural number. But $P^+(0, n)$ holds if n has all hereditary properties of 0 that all P-successors of 0 have. Hence, to define something as a natural number requires knowing if that something is a natural number, so the alleged circularity tells us. The question is whether this is a problem for Frege. The existence of circularity depends upon some previous philosophical background, and in Frege's realist perspective there seems to be none.

The other place where impredicativity affected Frege was in his full comprehension principle for functions. This, added to Axiom V regarding value-ranges, yields Russell's paradox. It was this problem that motivates Russell in avoiding such impredicative cases. Most of the important theorems regarding the Ancestral are dependent on comprehension axioms as well. Thus, in order to use them, it is necessary to provide some restrictions in order to avoid cases like $\exists F \forall x (F(x) \equiv \neg F(x))$ to be derivable[26]. But even with such restrictions, Frege's Ancestral works as intended. Other problems, and answers, that Frege's Ancestral is subject to are found in Heck (2016). In fact, some of the defense of Frege's definition here taken are found in Heck's paper, precisely on the misconception about quantification assumed by Kerry and Angelelli. What I proposed here, however, was an extended answer that relies more on

[26] In this case, by simply forbidding the formula F to occur free in the right side of the equivalence.

Frege's philosophy. To be historically fair, we can credit the early Russell of the *Principles of Mathematics* for being the first to point out the problem in such objections.

In conclusion, by taking a Fregean stance, and by arguing that Frege's definition of the Ancestral is not circular as both Kerry and Angelelli intended, we can reject the circularity premises in both arguments. In doing so, we are entitled to reject both conclusions, *viz.,* that Frege's definition is not verifiable, and that is not a proper reduction of the ordinary notion. With that, we kept Frege's reduction safe, at least with respect to his philosophical motivations, and showed that he could have been the first realist stance against vicious circle principles.

References

ANGELELLI, Ignacio. Frege's Ancestral and its Circularities. *Logica Universalis*, 6(3):477–483, 2012.

BOOLOS, George and HECK, Richard Kimberly. Die Grundlagen der Arithmetik §§82-83. In: HECK, Richard Kimberly, *Frege's Theorem*. Clarendon Press: Oxford, 2011. pp.69-89.

CARNAP, Rudolf. *Logical Syntax of Language*. Routledge University Press: London, 1937.

CARNAP, Rudolf. The Logicist Foundations of Mathematics. In: BENACERRAF, Paul; PUTNAM, Hilary (eds). *Philosophy of Mathematics:* Selected readings. 2ed. Cambridge: Cambridge University Press, 1964.

FEFERMAN, Solomon. Predicativity. In: SHAPIRO, Stewart (ed.) *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford University Press: New York, 2005. pp.590-624.

FREGE, Gottlob. *The Foundations of Arithmetic*. A logico-mathematical enquiry into the concept of number. Blackwell: Oxford, 1953.

FREGE, Gottlob. Begriffsschrift, a Formula Language, modeled upon that of Arithmetic, for Pure Thought. In: HEIJENOORT, Jean Van (ed.) *From Frege to Gödel*: A Source Book in Mathematical Logic, 1879-1931. Harvard University Press: Cambridge, 1967. pp.1-82.

FREGE, Gottlob. *Collected Papers on Mathematics, Logic and Philosophy*. Basil Blackwell: New York, 1984.

FREGE, Gottlob. *Basic Laws of Arithmetic*. Oxford University Press: Oxford, 2013.

FULUGONIO, Maria. Benno Kerry: una crítica temprana al logicismo de frege y un antecedente en la historia del problema de la impredicatividad. *CLE e-Prints*, 8(6), 2008.

GÖDEL, Kurt. Russell's Mathematical Logic. In: FEFERMAN, Solomon (ed.) Kurt Gödel Collected Works, volume II. Oxford University Press: Oxford, 1990. pp.119-141.

HECK, Richard Kimberly. Is Frege's definition of the Ancestral adequate? *Philosophia Mathematica*, 24(1):91–116, 2016.

KANT, Immanuel. *Critique of Pure Reason*. Cambridge University Press: Cambridge, 1998.

KERRY, Benno. Über Anschauung und ihre Psychische Verarbeitung. *Vierteljahrsschrift für Wissenschaftliche Philosophie*, III(11):249–307, 1887.

PARSONS, Charles. Realism and the Debate on Impredicativity, 1917–1944. In: PARSONS, Charles. *Philosophy of Mathematics in the Twentieth Century*: Selected Essays. Cambridge: Harvard University Press, 2014. pp.40-66.

POINCARÉ, Henri. *Mathematics and Science*: Last Essays. Dover Publications, Inc: New York, 1963.

RAMSEY, Frank. The Foundations of Mathematics. In: BRAITHWAITER, Richard (ed.). *The Foundations of Mathematics and other Logical Essays.* Routledge & Kegan Paul Ltd: London, 1931. pp.1-61.

RUSSELL, Bertrand. On some difficulties in the Theory of Transfinite Numbers and Order Types. *Proceedings of the London Mathematical Society*, s2-4(1):29–53, 1907.

RUSSELL, Bertrand. *The Principles of Mathematics*. W. W. Norton & Company, New York, 1996.

THE BAD COMPANY OBJECTION AND THE EXTENSIONALITY OF FREGE'S LOGIC

Vincenzo Ciccarelli[1]

ABSTRACT

According to the Bad Company objection, the fact that Frege's infamous Basic Law V instantiates the general definitional pattern of higher-order abstraction principles is a good reason to doubt the soundness of this sort of definitions. In this paper I argue against this objection by showing that the definitional pattern of abstraction principles – as extrapolated from §64 of Frege's *Grundlagen*– includes an additional requirement (which I call *the specificity condition*) that is not satisfied by the Basic Law V while is satisfied by other higher-order abstractions such as Hume's Principle. I also show that the failure of this additional requirement in the case of Basic Law V is engendered by an essential feature of Frege's conception of logic and thus that Frege himself should not have regarded the Basic Law V as a definition by abstraction.

Keywords: Abstraction Principles. Bad Company. Content Recarving. Frege's Logic.

RESUMO

Segundo a objeção da Má Companhia, o fato de que a infame Lei Básica V de Frege proporciona o padrão de definição geral dos princípios de abstração superior é uma boa razão para duvidar da validade deste tipo de definições. Neste artigo, eu argumento contra esta objeção, mostrando que o padrão de definição dos princípios de abstração — como extrapolados a partir do §64 do *Grunlagen* de Frege — inclui um requisito adicional (que denomino como *a condição de especificidade*) que não é satisfeito pela Lei Básica V, embora seja satisfeito por outras abstrações de ordem superior, tal como o Princípio de Hume. Mostro também que a falha deste requisito adicional no caso da Lei Básica V é engendrada por uma característica essencial da concepção de Frege da lógica, e que, assim, o próprio Frege não deveria ter tomado a Lei Básica V como uma definição por abstração.

Palavras-chave: Princípios de Abstração. Má Companhia. Reentalhamento do Conteúdo. Lógica de Frege.

[1] Universidade de São Paulo
Email: ciccarelli.vin@gmail.com. ORCID: 0000-0002-8996-9170.

1. Introduction

Frege's early attempt to provide logical foundations to arithmetic is based on a special sort of definitions which have been named *abstraction principles*. Although Frege neither used this terminology nor explicitly formulated a rigorous condition on what it takes for a definition to be an abstraction principle, it is commonly agreed that an abstraction principle is an implicit definition of instances of a certain concept, where conditions of identity between them are framed in terms of a an equivalence relation. One of Frege's most famous examples of an abstraction principle is the implicit definition of the concept of 'direction of a straight line': the direction of a line *a* is identical to the direction of a line *b* iff line *a* is parallel to line *b*. In general, the definitional pattern of a first-order abstraction principle is formally represented as follows:

$$\forall x \forall y[R(x, y) \leftrightarrow \sigma(x) = \sigma(y)] \quad (\mathbf{AP})$$

where R is a first-order equivalence relation and σ is the so-called abstraction operator. It has to be said that abstraction principles which play an effective role in Frege's logicist project are second-order, i.e. the equivalence relation appearing on the left limb of the bi-conditional is second-order. For instance, *Hume's Principle* – i.e. the implicit definition of cardinal numbers – is framed in terms of the relation of *equinumerousity:* the concept *F* is equinumerous to the concept *G* iff the cardinal number of *F* is identical to the cardinal number of *G*. We schematically represent a second-order abstraction principle as follows:

$$\forall X \, \forall Y \, [\mathbf{R}_x(Xx, Yx) \leftrightarrow \sigma_x(Xx) = \sigma_x(Yx)] \qquad (\mathbf{AP2})$$

where X, Y are second-order variables, **R** is a second-order equivalence relation, and σ is a second-order abstraction operator (an operator that when combined with a monadic conceptual expression forms a singular term). Another famous (and notorious) example of higher-order abstraction is Frege's Basic

Law V which says that whenever two concepts are co-extensional they have the same *value-range*:

$$\forall X\, \forall Y\, [\forall x(Xx \leftrightarrow Yx) \leftrightarrow \varepsilon_x(Xx) = \varepsilon_x(Yx)] \qquad (\mathbf{BLV})$$

where ε is the value-range operator. It is well-known that the Basic Law V leads to Russell's paradox and thus is inconsistent. This fact has induced a general objection against the soundness of definitions by abstraction which has been traditionally called *the Bad Company objection*. The objection may be summarized as follows: if the definitional pattern of abstraction principles only requires that the relation appearing in the left limb is an equivalence relation, then this definitional pattern cannot be the base for formulating correct definitions, for the relation of co-extensionality appearing in the left limb of the Basic Law V is an equivalence relation, yet the correspondent abstraction principle is inconsistent. Thus abstraction principles that are reasonably consistent – such as Hume's Principle – are in bad company, for they share the same definitional pattern with the Basic Law V.

As remarked by Linnebo (2009) "The bad company problem shows that a deeper understanding is needed of the conditions under which abstraction is permissible". In other words, the fact that a relation *R* is an equivalence relation is not enough to ensure that the correspondent bi-conditional (**AP2**) should count as an abstraction principle, i.e. as a valid definition. A possible reply to the bad company objection is based on a deeper analysis of the definitional pattern that characterizes abstraction principles resulting in new requirements on the relation appearing in the left limb such that the Basic Law V may be shown to fail to satisfy these requirements that other reasonably consistent abstraction principles meet. This is the strategy that I pursue in this paper. In section 2 I will analyze the definitional pattern of abstraction principles by considering the passage in which Frege introduces this sort of definitions; I will argue that in order to formulate a permissible abstraction principle, the equivalence relation *R* must satisfy what I call *the specificity condition*. In section 3 I will show that differently from Hume's Principle, the Basic Law V do-

es not satisfy the specificity condition. In section 4 I will try to spell out what feature of the relation of co-extensionality in Frege's system of logic prevents the Basic Law V to be a permissible abstraction principle.

2. The definitional pattern of abstraction principles

The notion of definition by abstraction is introduced by Frege in §64 of his *Grundlagen*, where he introduces the famous example of the implicit definition of the concept of 'direction of a straight line':

> The judgment '*line a is parallel to line b*', [...] can be taken as an identity. If we do this, we obtain the concept of direction and say: '*The direction of line a is identical with the direction of line b*'. Thus we replace the symbol ∥ by the more generic symbol =, through removing what is specific in the content of the former and dividing it between *a* and *b*. We carve up the content in a way different from the original way, and this yields us a new concept. (FREGE, 1950), pp. 74-75

The main difficulty concerning this passage consists in the understanding of the crucial notion of 'content' (*Inhalt*) and the way this notion is handled. For instance, it is not clear how to understand the fact that a certain part of the content of the relation of parallelism is "removed" and successively "divided" between *a* and *b*. What seems to be clear from the passage is that the sentence 'the direction of *a* is identical to the direction of *b*' may be obtained by performing certain content preserving operations on the sentence 'line *a* is parallel to line *b*'; as a result, the two sentences express the same content. To my knowledge, the relevant literature has focused mainly on attempts to define such a relation of identity of content between sentences ((HALE, 1997), (POTTER; SMILEY, 2001), (YABLO, 2008)) rather than on spelling out the operation of content removal and division described in the passage. Albeit this paper does not aim at entirely explaining away Frege's metaphorical language, some hints on the definitional pattern Frege is trying to present may be extrapolated.

I understand the passage as saying that the same content may admit different internal organizations that result in different syntactic structures. For instance, being *C* of the sentence '*a* ∥ *b*', we may obtain a different internal or-

ganization of *C* by performing the procedure described in the passage on *C* as already organized according to the syntactic structure of '*a* ∥ *b*'. In particular, we may remove part of the content of the relation of parallelism and attach it – in an unspecified way – to the content of '*a*' and '*b*'. As a result, the "remainder" of the content of the relation of parallelism after removing its "specific content" is the content of the relation of identity; and the outcome of attaching this content to '*a*' and '*b*' is the content of new singular terms, i.e. 'the direction of *a*' and 'the direction of *b*'. Thus the two limbs of an abstraction principle have the same content in virtue of the fact that what is removed from one constituent of the content is attached to others.

As mentioned in the previous section, there is a general agreement on the fact that the relation *R* appearing in the left limb of an abstraction principle must be an equivalence relation. Yet Frege seems to suggest that something more is required, something that at first glance may appear almost trivial, but as we will see is a fundamental requirement to rule out some abstraction principles as unsound. Frege explicitly appeals to a specific part of the content of an equivalence relation; moreover, when he says that we replace a relational symbol "by the more generic symbol" of identity, he seems to suggest that what is specific in the content of an equivalence relation *R* is what makes it more specific than identity, i.e. what characterizes the difference in content between *R* and =. In what sense does an equivalence relation *R* is more specific than an identity (i.e. has a content including a specific part with respect to the relation of identity)? Frege does not give any hint regarding such a comparison between equivalence relations. Thus a hypothetical answer to the previous question must be formulated outside of the context of Frege's texts.

An interesting way of understanding the higher specificity of the relation of parallelism with respect to identity may come from some fundamental algebraic properties of equivalence relations. The idea is the following: an equivalence relation is a congruence with respect to a restricted range of properties – commonly referred to as *invariant properties* – while identity is a congruence with respect to all properties. In other words, two parallel lines are

indistinguishable with respect to their *orientational properties*, whereas two identical lines are indistinguishable with respect to all properties. As a consequence, an equivalence relation R may be understood as a relation of *partial or restricted indistinguishability* while an identity is understood as a relation of total or general indistinguishability. Given an equivalence relation R, the more restricted is the range of R-invariant properties, the more specific R is with respect to the relation of identity. This higher specificity of an equivalence relation R may be formalized in terms of a restricted quantification over properties. Being I_R the higher-order predicate 'X is an R-invariant property', the following holds for every two individuals a and b:

$$Rab \rightarrow \forall X\,[I_R(X) \rightarrow (Xa \leftrightarrow Xb)] \qquad (*)$$
$$a = b \rightarrow \forall X\,[Xa \leftrightarrow Xb] \qquad (**)$$

In other words, the relation R is concerned with a specific range of properties I_R, whereas the relation of identity is concerned with all possible properties that individuals may have. The formulas $(*)$ and $(**)$ may be used to formulate a crucial requirement that a relation R must satisfy in order to run Frege's principle of content recarving. According to the quoted passage of §64, the content of an abstraction operator is formed using the specific part of the content of an equivalence relation R and when we remove this part from the content of R we obtain the content of the relation of identity. Whatever expressions like "removing" and "dividing" a content might mean, it is clear that if there is no "specific part" of the content of R, there is no available content to be "divided" to define the content of the abstraction operator. Therefore an equivalence relation R must satisfy the following condition: *R must express a content that includes "a specific part"*. According to the previous remark, a relation R expresses a content including a specific part only if R is more specific than the relation of identity, which in turn may be understood as the fact that R admits a range of invariant properties that does not coincide with the totality of properties. In other words, to determine whether R is more specific than identity we have to check whether there is a property φ that is

not an *R*-invariant property. How may this requirement on properties be expressed? According to (∗) if a property φ is such that given two *R*-related individuals *a* and *b*, either a has φ and *b* lacks φ, or *a* lacks and *b* has φ. In other words, if *Rab* is true and a property φ may be used to distinguish between *a* and *b*, then φ is not an *R*-invariant property, i.e. *R* is more specific than identity. We may summarize this reasoning by stating the following requirement:

> **Specificity condition (first-order)**
> *A first-order equivalence relation R expresses a content including a specific part iff there are two individuals a, b and a property* φ *such that:*
>
> $Rab \wedge [(\varphi(a) \wedge \neg\varphi(b)) \vee (\varphi(b) \wedge \neg\varphi(a))]$ (**SC**)

The specificity condition may appear as a trivial requirement: it suffices that an equivalence relation *R* is not the relation of identity for *R* to satisfy (**SC**). It is immediately evident that the relation of parallelism satisfies this requirement, for surely there are two distinct straight lines *a* and *b* that are parallel. As we will see, the specificity requirement becomes interesting when we consider higher-order equivalence relations:

> **Specificity condition (higher-order)**
> *A second-order equivalence relation* ***R*** *expresses a content including a specific part iff there are two concepts F, G and a second-order property* Φ *such that:*
>
> $\boldsymbol{R}_x(Fx, Gx) \wedge [(\Phi_x(Fx) \wedge \neg\Phi_x(Gx)) \vee (\Phi_x(Gx) \wedge \neg\Phi_x(Fx))]$ (**SCHO**)

In this case it makes no sense to say that ***R*** meets the specificity condition whenever is distinct from identity, for there is no defined identity relation between concepts: identity is characteristic of objecthood. However, even though there is no relation of identity between concepts, it is possible to define a relation of distinguishability which is expressed by the fact that there is a se-

cond-property property that one concept has and the other lacks. It is evident that the relation of equinumerousity satisfies the specificity condition; let *F* be the concept defined by the open formula $x = a$ and *G* defined by $x = b$. Clearly, *F* and *G* are equinumerous (both have only one instance). The second-order property '*X* has *a* as instance' is a property that *F* has and *G* lacks, thus the property '*X* has *a* as instance' may be used to distinguish between two equinumerous concepts. The main result of this section is that the definitional pattern of an abstraction principle is not limited to the fact that the relation appearing in the right-limb must be an equivalence relation, for the specificity condition must also be met in order to run the procedure of content recarving and provide a content to the abstraction operator. In the next section we will see that the relation of co-extensionality – which appears in the right limb of the Basic Law V – does not meet this requirement. This is a crucial point to argue that the Basic Law V is not a bad company to Hume's Principle, for – in virtue of the fact that it does not instantiate the same definitional pattern – it is not even a company.

3. Basic Law V

As anticipated at the end of the previous section, in this section I will show that the relation of co-extensionality does not satisfy the specificity condition in both Frege's system of logic and standard higher-order logic, and thus cannot be conceived as an instance of the definitional pattern of abstraction principles. The main consequence of this fact is that the Basic Law V cannot be considered as a "bad company" for Hume's Principle. As a preliminary historical remark, notice that the view that according to Frege the Basic Law V should be obtained by recarving the content of the relation of co-extensionality is not uncontroversial. Frege never justifies the truth of this axiom by invoking the procedure of §64; he rather seems to consider the Basic Law V as a primitive logical law[2]. The fact that in the *Grundgesetze* he considers the two limbs of the Basic Law V as being "gleichbedeutend" (i.e. identical in meaning) is no evidence for the thesis that they should be identical in content: for

[2] (FREGE, 2013, p.14)

at the time the *Grundgesetze* was written, the word "gleichbedeutend" is undoubtedly translated as "identical in reference", thus identical in truth value. As a consequence, nothing in Frege's mature writings suggests that he ascribed a relation between the two limbs of the Basic Law V stronger than logical equivalence. However, in the essay *Function and Concept* (FREGE, 1952) Frege seems to suggest that a sentence saying that two concepts are co-extensional expresses the same sense as a sentence saying that their extensions coincide (as highlighted in (BURGE, 1984)); hence, if we assume that identity of sense implies identity of content – i.e. that the notion of content is coarser in grain than the notion of sense – then Frege's example suggests that the two limbs of the Basic Law V are identical in content. And this fact represents a good reason to suppose that an identity of extensions is obtained by a content recarving procedure from the content of the relation of co-extensionality. The main purpose of this paper is to show that the Basic Law V cannot be considered as a bad company of Hume's Principle due to the fact that it does not satisfy the definitional pattern described in §64 of the *Foundations*. Thus it is not crucial for our purpose to show that Frege should have considered the Basic Law V and Hume's Principle as different instances of the same definitional pattern.

In the remaining part of this section I will show that the Basic Law V cannot be considered as an instance of the definitional pattern described in *Grundlagen* §64 in virtue of the fact that it fails to satisfy the specificity condition. We will start by formulating this condition for the relation of co-extensionality, i.e. the equivalence relation that appears in the left limb of the Basic Law V:

> **Specificity Condition (BLV)**
>
> *The relation of co-extensionality expresses a content having a specific part iff there are two concepts F, G and a second-order concept* Φ *such that:*

$$\forall x(Fx \leftrightarrow Gx) \wedge [\Phi_x(Fx) \wedge \neg\Phi_x(Gx)] \vee [\Phi_x(Gx) \wedge \neg\Phi_x(Fx)]$$

(**SCBLV**)

The fact that the relation of co-extensionality does not satisfy (**SC-BLV**) in the formal system of the *Grundgesetze* may be proven by showing that the following principle – which I call the *Principle of Extensional Equivalence* – holds: for every higher-order open formula $\Phi_x(Xx)$ where the variable X occurs free,

$$\forall x(Fx \leftrightarrow Gx) \rightarrow (\Phi_x(Fx) \rightarrow \Phi_x(Gx)) \qquad (\textbf{PEE})$$

The Principle of Extensional Equivalence says that two co-extensional concepts are interchangeable *salva veritate* in all contexts. In other words, there is no higher-order formula that may distinguish between two co-extensional concepts. The proof of (**PEE**) in Frege's system will be presented through an adaptation of Frege's notation to make it understandable even to the reader who is not familiar with the concept script. Let ι be an operator that when applied to a value range v returns the value c if c is the only element of v, otherwise the false. Following §34 of (FREGE, 2013) we define the membership operator as follows:

$$(a \in b) = \iota\, \varepsilon_z[\exists g\, (b = \varepsilon_u g(u) \wedge g(a) = z)] \qquad (\textbf{DEF})$$

which says that the function $a \in b$ of the arguments a, b has the same value as the function g of the argument a if there is a function g such that the value range of g is b, otherwise the value of $a \in b$ is the False.

Let ξ be a metavariable and f a first-order function of one argument, by considering the following instance of (**DEF**):

$$(\xi \in \varepsilon_u f\,(u)) = \iota\, \varepsilon_z\,[\exists g\, (\varepsilon_u f\,(u) = \varepsilon_u\, g(u) \wedge g(\xi) = z)]$$

and by applying the Basic Law V, it follows that:

$\xi \in \varepsilon_u f(u) = f(\xi)$ (**T1**)

In other words, for every function f and every term ξ, the expression '$\xi \in \varepsilon_u f(u)$' is coreferential to the expression '$f(\xi)$'. We can now prove the Principle of Extensional Equivalence. Let $A(f(\xi))$ a formula in which the function f occurs applied to a term ξ (ξ may be a Roman, German, or uppercase Greek letter). A sketch of the proof follows:

(7) $A(f(\xi))$	(Assumption)
(8) $\forall x(f(x) = g(x))$	(Assumption)
(9) $A(\xi \in \varepsilon_u f(u))$	(by (**T1**), 1, and Leibniz's Law)
(10) $\varepsilon_u f(u) = \varepsilon_u g(u)$	(by Basic Law V, 2, and Modus Ponens)
(11) $A(\xi \in \varepsilon_u g(u))$	(by Leibniz's Law on 3 and 4)
(12) $A(g(\xi))$	(by (**T1**), 5, and Leibniz's Law)

Which proves that for every formula A in which the function f occurs, if $A(f(\xi))$ and if f and g are co-extensional, then $A(g(\xi))$. Therefore, in Frege's system co-extensional functions are always replaceable *salva veritate*. Hence there is no higher-order formula of Frege's system that may distinguish between co-extensional functions which, in turn, means that the relation of co-extensionality does not satisfy the specificity condition. Considered that – as argued in the previous section – the specificity condition should be included as a requirement that an equivalence relation must satisfy in order to run the content recarving procedure and thus introducing the corresponding abstraction principle, the definitional pattern of abstraction principles extrapolated from §64 of the *Grundlagen* is not applicable to the relation of co-extensionality. As a consequence, the Basic Law V and Hume's Principle are not instances of the same definitional pattern, i.e. the Basic Law V is not a bad company for Hume's Principle, for it is not even a company. Notice that the fact that the relation of co-extensionality does not satisfy the specificity condition implies that it expresses no specific content in Frege's logic, thus there is no available content to form the value range operator.

Contemporary higher-order logic deserves some additional comments. The formula (**PEE**) is valid in both standard and Henkin's semantics for higher-order logic; given the incompleteness of higher-order logic, we cannot take this fact as a proof of the fact that the formal system treats conceptual expressions in a pure extensional way. However, as remarked in (HECK, 2012) (p. 137), the principle of extensional equivalence may be proven by induction in the complexity of higher-order formulas. Therefore, both Frege's logic and standard higher-order logic are *purely extensional*, i.e. they allow for the replacement of co-extensional conceptual expressions in every formula *salva veritate*.

One may object that the proposed argument is circular when applied to Frege's system, for – as previously mentioned – we use the Basic Law V to prove the principle of extensional equivalence. Notice that to show that the relation of parallelism satisfies the specificity condition we did not need to assume the abstraction principle of directions; similarly, to show the same fact regarding the relation of equinumerousity, we did not need to assume Hume's Principle.

Such a circularity may induce a new version of the bad company objection: given that the Basic Law V is false (due to Russell's paradox), we cannot conclude that (**PEE**) is valid in Frege's system of logic. As a consequence, the relation of co-extensionality may still satisfy the specificity condition and thus the Basic Law V may still be a company for Hume's Principle; in particular, given the falsity of the Basic Law V, it will be a bad company. The objection may be replied by remarking that the Basic Law V may be dispensable in the proof of (**PEE**); indeed, it is very plausible to assume that we may adapt the proof of (**PEE**) in standard higher-order logic described by Heck (2012) to Frege's system. In the next section we will try to understand in further details which feature of both Frege's logic and Frege's conception of higher-order predication prevents the relation of co-extensionality to satisfy the specificity condition.

4. Frege's logic and higher-order predication

In this section we will try to understand more deeply in virtue of which characteristic of Frege's logical system the relation of co-extensionality fails to satisfy the specificity condition. We have seen that the relation of co-extensionality fails to satisfy the specificity condition due to the fact that the principle of extensional equivalence is provable in Frege's logical system. Thus our question becomes: how the proof of the principle of extensional equivalence in Frege's logical system may be blocked? The proof of the principle of extensional equivalence uses basic features of higher-order logic, such as Leibniz's Law and *Modus Ponens* whose validity in the present context is considered out of question. The only point that is debatable is the theorem labeled (**T1**), which says that for every first-order function f and argument ξ, the expression '$f(\xi)$' may always be replaceable by '$\xi \in \varepsilon_u f(u)$' *salva veritate*. It is important to notice that (**T1**) does not says that the incomplete expressions '$f(...)$' and '$... \in \varepsilon_u f(u)$' are names of the same function, yet names of different co-extensional functions. Nevertheless, if '$f(...)$' and '$... \in \varepsilon_u f(u)$' denote different functions, it should be in principles possible to assert something true of one and false of the other, thus making the principle of extensional equivalence fail. In informal examples, we use free variables to substantivize functions, as in the expression 'the concept *x is a horse*'. And this makes possible to construct examples of sentences in which the intensional difference between functions implies an extensional difference between sentences: the sentence 'John grasps the concept *x is a horse*' may not have the same truth-value as the sentence 'John grasps the concept *x is in the value range of the concept 'y is a horse'*', perhaps by assuming that John grasps the concept horse while being innocent of the notion of value range. Hence the possibility of expressing contexts in which the principle of extensional equivalence fails relies on the possibility of substantivizing functions.

In Frege's concept script there is no way of substantivizing functions. More precisely, Frege does not use predicate letters that can be used as function names to form higher-order predications directly applied to functions; moreover, there are no symbols of the language denoting empty argument places;

not even free variables, given that even Roman letters used to express generality have a defined (implicit) scope, thus acting as bound variables. When Frege wants to denote a function in his informal comments, he either uses meta-variables such as 'ξ' or dots standing for empty argument places. This point may be clarified a bit further. Given a first-order function f, I define *a direct higher-order predication of f* as a formula in which f occurs with no specification of its argument or using a free variable as argument. For instance '*square root* is a continuous function' or 'the concept *x is a horse* is difficult to grasp' are cases of direct higher order predication. On the other hand, I define *an indirect higher-order predication on f* as a formula in which f occurs with its argument placed filled by a proper name of an object, a bounded variable, or a variable with a restricted scope. In Frege's logic only indirect higher-order predication is possible, for there are no letters denoting function names which may be used without argument specification and the argument places of first-order functions may be filled either by uppercase Greek letters (proper names of objects) or by German letters (bounded variables) or by Roman letters (variables with restricted scopes)[3].

The impossibility of substantivizing functions – and thus of direct higher-order predication – is not something that Frege overlooked or that escaped his attention; it is considered as a desirable feature of his logical system, for it allows for *the reduction of orders*, as he clearly says:

> First-order functions can be used instead of second-level functions in what follows. This will now be shown. As was indicated, this is made possible by the fact that the functions appearing as arguments of second-level functions are represented by their value-ranges, although of course not in such a way that they simply concede their places to them, for that is impossible. (FREGE, 2013, p.52)

In spite of the fact that Frege recognizes that we cannot simply replace a function name by the name of its value range, Frege holds that value ranges shall be considered as *representative objects* of concepts in higher-order predi-

[3] There is also the case of lowercase Greek letters used to represent schematically higher-order functions, as in the expression $M_\beta(\, f\,(\beta)\,)$. However, the presence of the index β in M_β clearly suggests that higher-order functions act as variable binders, thus allowing the substitution $M_\beta(\, \beta \in \varepsilon_u\, f\,(u)\,)$.

cation[4]; as a consequence, a higher-order predication applied to a concept $f(x)$ may always be converted into a first-order one applied to the value range of $f(x)$.

This leads to the role of Basic Law V in Frege's logic and to the main point of this paper. As shown by Wright (1983), Frege could have done without the Basic Law V and derive the axioms of arithmetic directly from Hume's Principle plus standard higher-order logic. However, Frege did not regard the possibility of deriving the axioms of arithmetic and of defining numbers as logical objects as the only role of value ranges in his logical system. As remarked by Heck (2012) and as shown by the previous quotation, Frege regarded the Basic Law V as having the virtue of *lowering the orders*, i.e. of converting a higher-order predication into a first-order one by introducing value ranges as representative objects of functions. Hence, the additional and crucial role of the Basic Law V is that of making the logic purely extensional by reducing assertion about concepts to assertions about their value ranges. In other words, Frege regarded the Basic Law V as the principle that makes explicit that co-extensional concepts are replaceable *salva veritate* in all contexts. I have shown that this amounts to say that the Basic Law V does not satisfy the specificity condition which implies that it is not a bad company for Hume's Principle. Therefore, Frege's conception of logic requires that the Basic Law V is not an abstraction principle *on a par* with the others; the special role he ascribes to the Basic Law V implies that the relation of co-extensionality must be a sort of identity between concepts, i.e. an absolutely general higher-order equivalence relation with no specific content. Thus the fact that the Basic Law V does not instantiate the definitional pattern described in §64 of Grundlagen is not an unexpected consequence of Frege's axiomatization, yet is essential to the role that value ranges play in Frege's logic.

5. Conclusion

In this paper I have proposed an argument showing that the bad company objection may not be as harmful as it is commonly thought. More precisely, I have shown that the objection relies on the false assumption that the

[4] (RUFFINO, 2000), (RUFFINO, 2003)

Basic Law V is an instance of the definitional pattern behind all abstraction principles. My argument is based on the assumption that the definitional pattern of abstraction principles is not characterized just by the requirement that the relation appearing in the left limb is an equivalence relation. By analyzing the passage in §64 of the *Grundlagen*, I have shown that a further condition must be met, i.e. what I have called *the specificity condition*. Notice that this is a particular view on what abstraction principles are: one may consider abstraction principles from a different perspective, i.e. independently of how Frege introduces and justifies them, and thus may not accept a requirement extrapolated from §64 of the *Grundlagen* as a necessary condition for the truth of an abstraction principle. Moreover, I have argued that, according to Frege's conception of logic and to his view on the role of value ranges of functions, it is essential to the possibility of reducing the orders that the relation of co-extensionality does not satisfy the specificity condition and thus that the Basic Law V shall not instantiate the definitional pattern extrapolated from *Grundlagen* §64. It is worth clarifying that this paper is not directly concerned with the problem of the inconsistency of the Basic Law V. I have not tried to argue that the Basic Law V is inconsistent *in virtue of* the fact that in Frege's system of logic the relation of co-extensionality does not satisfy the specificity condition. Another point that is worth highlighting is that the argument I have presented is not aimed at formulating a complete list of requirements that abstraction principle must satisfy in order to be acceptable. In other words, I am not saying that given an equivalence relation *R*, if *R* satisfies the specificity condition, then the corresponding abstraction principle is true. There might be further requirements that the proposed analysis fails to capture. As a consequence, this paper has mainly a negative aim: it shows that being the specificity condition a necessary condition, and given that the Basic Law V fails to satisfy this condition, then it cannot be counted as an instance of the same definitional pattern of Hume's Principle. However, I presented no argument to the effect that the specificity condition is also sufficient; hence, there may be other inconsistent abstraction principles that satisfy the specificity condition,

thus showing that a deeper analysis of the definitional pattern of abstraction principles is still needed.

On the other hand, I have presented an interesting alternative approach to the philosophical issues regarding abstraction principles by turning more attention on the way Frege introduces this sort of definitions and the justification for their truth that he seems to suggest based on the procedure of content recarving.

References

BURGE, T. Frege on Extensions of Concepts, From 1884 to 1903. **Philosophical Review**, Durham, v. 93, n. 1, p. 3-34, 1984.

FREGE, G. **Foundations of Arithmetic**. Routledge, 1950

FREGE, G. Translations From the Philosophical Writings of Gottlob Frege. In GEACH P, **Early Analytic Philosophy – New Perspectives and Tradition**. Blackwell, 1952

FREGE, G. **Gottlob Frege: Basic Laws of Arithmetic**. Oxford University Press, 2013

HALE, B. Grundlagen §64. **Proceedings Of The Aristotelian Society, New Series**, Londres, v. 97, p. 243-261, 1997

HECK, R. G. **Reading Frege's Grundgesetze**. Oxford University Press, 2012

LINNEBO, O. Bad Company Tamed. **Synthese**, v. 170, n. 3, p. 371-391, 2009

POTTER, M.; SMILEY, T. Abstraction by Recarving. **Proceedings Of The Aristotelian Society**, Durham, v. 101, n. 3, p. 327-338, 2001.

RUFFINO, M. Extensions as Representative Objects in Frege's Logic. **Erkenntnis**, v. 52, n. 2, p. 239-252, 2000.

RUFFINO, Marco. Why Frege would not be a Neo-Fregean. **Mind**, v. 112, n. 445, p. 51-78, 2003.

WRIGHT, C. **Frege's Conception of Numbers as Objects.** Aberdeen University Press, 1983.

YABLO, S. Carving the Content at its Joints. **Canadian Journal of Philosophy**, n.38, p.145- 177, 2008.

QUANTIFICADORES E O PROBLEMA DE FREGE

Bruno Ramos Mendonça[1]

RESUMO

Em *Sobre o Sentido e a Referência*, Frege (2009) propôs a seguinte questão (no que segue, chamarei essa questão de 'problema de Frege'): como é possível que frequentemente ampliemos nosso conhecimento ao descobrir a verdade de identidades da forma $a=b$? A esse problema, um caso particular do problema da onisciência lógica, Frege dá a seguinte resposta em termos da distinção sentido-referência: alguém que ignora a validade de $a=b$ não sabe que os sentidos de ambas as expressões denotativas envolvidas na equação identificam a mesma referência. Apesar de não ser consensual, essa solução é bastante influente. Logo, seria interessante verificar se ela poderia ser generalizada para outros casos do problema da onisciência lógica. O presente artigo aborda esse tema. Pretendo aqui mostrar que o problema de Frege e a sua solução em termos da distinção sentido-referência são generalizáveis a uma família mais ampla de casos de falha de onisciência lógica associados à nossa competência semântica no uso de quantificadores. Enquanto o problema de Frege original trata de casos de ignorância sobre a correferencialidade de dois termos singulares, a versão generalizada do problema trata de casos de ignorância sobre a co-extensionalidade de dois ou mais quantificadores ocorrendo em uma fórmula. Apelando a uma caracterização bidimensionalista do conceito de sentido, sustentarei que os quantificadores apresentam um elemento indexical e, consequentemente, estão associados a uma função caráter que, em certos contextos linguísticos tolerados pela situação epistêmica imperfeita de agentes racionais médios, atribuem domínios não co-extensionais a diferentes quantificadores ocorrendo em uma fórmula.

Palavras-chave: Problema de Frege. Quantificação. Onisciência Lógica. Sentido e Referência. Semântica Bidimensional.

ABSTRACT

In *On Sense and Reference*, Frege (2009) tackles the following question (hereafter, I refer to it simply as 'Frege's puzzle'): how can we often augment our knowledge by discovering the truth of identities of the form $a=b$? To this question, a subcase of the logical omniscience problem, Frege offers the

[1] Pesquisador de pós-doutorado CAPES (#88887.475401/2020-00) no Centro de Lógica, Epistemologia e História da Ciência da Universidade Estadual de Campinas (CLE/ Unicamp).
E-mail: bruno.ramos.mendonca@gmail.com. ORCID: 0000-0002-0583-4042.

following response in terms of a sense and reference dichotomy: an individual who ignores that *a=b* is true does not know that the senses of the proper names occurring in the sentence pick out the same reference. Although not uncontroversial, Frege's solution is still very influential. Hence, it would be interesting to verify whether it can be generalized for other cases of the logical omniscience problem. The present paper deals with this issue. I want to show that Frege's puzzle as well as its solution in terms of a distinction between sense and reference can be applied to a broader set of cases of failure of logical omniscience related to our semantic competence in using quantifiers. Whereas Frege's original puzzle considers cases of ignorance on the correferentiality of two proper names, the generalized version of the puzzle regards cases of ignorance on the co-extensionality of two or more quantifiers occurring in a formula. Drawing upon a 2-dimensionalist characterization of the concept of sense, I hold that quantifiers display an indexical feature and, consequently, are associated to character functions that, in some counterfactual linguistic contexts not ruled out by the imperfect epistemic situation of ordinary rational agents, ascribe non-extensional domains to different quantifiers occurring in the considered formula.

Keywords: Frege's Puzzle. Quantification. Logical Omniscience. Sense and Reference. 2-Dimensional Semantics.

1. Introdução

Em *Sobre o Sentido e a Referência*, Frege (2009) propôs a seguinte questão (no que segue, chamarei essa questão de 'problema de Frege' simplesmente): como é possível que frequentemente ampliemos nosso conhecimento ao descobrir a verdade de identidades da forma *a=b*? Para ficarmos em um exemplo fregeano, por muito tempo os astrônomos não sabiam que os nomes 'Hesperus' e 'Phosphorus' denotam o mesmo corpo celeste. A descoberta desse fato aumentou o nosso conhecimento astronômico. No entanto, a identidade de algo consigo mesmo é um propriedade lógica fundamental de todos os entes. Logo, cabe perguntar, como é possível que nós, agentes racionais, por vezes ignoremos fatos lógicos tão básicos?

Na literatura contemporânea, uma versão mais geral desse problema (o assim chamado *problema da onisciência lógica*) tem interessado aos estudiosos das lógicas epistêmicas, lógicas modais que formalizam a expressão 'S sabe que...' em termos de um operador de necessidade. Os sistemas padrão de lógica epistêmica são lógicas modais *normais*, i.e., sistemas lógicos em que

vale o axioma K ('$K(p \rightarrow q) \rightarrow (Kp \rightarrow Kq)$') e a regra de necessitação: $p \Rightarrow Kp$. Ora, da normalidade desses sistemas formais se segue a propriedade de onisciência lógica: se S sabe que um conjunto de hipóteses T é o caso e q é uma consequência lógica de T, então S sabe que q é o caso. No limite, dado que as verdades lógicas são implicadas por qualquer conjunto de premissas, para toda verdade lógica q, S sabe que q é o caso (HENDRICKS & SYMONS, 2015, sec. 1). Contudo, a onisciência lógica é uma propriedade irrealista das lógicas epistêmicas normais. Um agente racional médio não conhece de antemão todas as verdades lógicas. Nesse sentido, há um interesse na literatura recente por definir sistemas de lógica epistêmica não-normais que validem somente versões enfraquecidas da onisciência lógica.

O problema da onisciência lógica está longe de ser apenas uma questão técnica sobre sistemas formais. Há uma contraparte conceitual do problema cuja solução não é óbvia: se um agente racional médio não conhece todas as verdades lógicas, quais são os fatos lógicos que ele eventualmente desconhece e quais são as razões subjacentes a essa ignorância? Sem uma análise prévia dessa questão, não podemos decidir qual enfraquecimento da propriedade de onisciência lógica é o mais adequado.

No seu ensaio clássico, Frege dá uma resposta à contraparte conceitual do problema da onisciência lógica restrito ao caso da identidade lógica. Resumidamente, Frege sustenta que o conteúdo semântico de expressões denotativas tais como os nomes próprios deve ser analisado em duas partes: se, por um lado, uma tal expressão possui uma referência, que é o ente por ela denotado, por outro lado essa relação de denotação é sempre mediada pelo sentido da expressão, i.e., o "modo de apresentação" (FREGE, 2009, p. 131) da referência. Embora seja tema de disputa entre os estudiosos da obra fregeana o que exatamente se deve entender pelo sentido de uma expressão denotativa, em termos gerais, com esse conceito Frege pretende capturar algum dispositivo semântico com base no qual os falantes podem identificar a referência de uma expressão denotativa. Assim, por exemplo, poder-se-ia argumentar que o sentido de um nome próprio (e.g., 'Platão') é uma propriedade que, em certo contexto de uso da linguagem, é satisfeita apenas pela referên-

cia do nome em questão (e.g., a propriedade 'ter sido professor de Aristóteles' poderia cumprir o papel de sentido do nome 'Platão').

Segundo o enquadramento fregeano, a competência semântica de um falante no uso de expressões denotativas está associada ao conhecimento do sentido dessas expressões. De fato, por vezes é possível que o falante até mesmo ignore se um dado nome próprio possui referência. Além disso, uma série de nomes próprios de nossa linguagem são referencialmente vazios, mas ainda assim são semanticamente carregados na medida em que possuem sentido. Com base na distinção sentido-referência, Frege explica o fenômeno da ignorância sobre o valor de verdade de identidades da forma $a=b$ nos seguintes termos: um sujeito que ignora a validade de $a=b$ não sabe que os sentidos de ambas as expressões denotativas envolvidas na equação identificam a mesma referência.

Ainda que esteja longe de ser consensual, a solução do problema de Frege em termos da distinção sentido-referência é bastante influente. Logo, seria interessante verificar se seria possível generalizar essa solução para outros casos do problema da onisciência lógica. O presente artigo aborda esse tema. Pretendo aqui mostrar que o problema de Frege (e a sua solução em termos da distinção sentido-referência) é generalizável a uma família mais ampla de casos de falha de onisciência lógica associados à nossa competência semântica no uso de quantificadores.

Os quantificadores são expressões de nosso vocabulário lógico de difícil categorização. Em primeiro lugar, deve-se notar que os quantificadores não são simples conectivos lógicos. Os conectivos conjunção, negação etc. expressam funções de verdade, mas os quantificadores apenas com um alto custo teórico poderiam ser assim concebidos (FREIRE, 2015). Portanto, se pretendemos evitar a conclusão de que os quantificadores configuram uma espécie *sui generis* de expressão linguística, nos restam apenas duas alternativas de análise: ou os quantificadores constituem uma forma de predicação, ou eles são um tipo especial de expressão denotativa.

Frege (1983) foi um proponente importante da ideia de que quantificadores são predicados de segunda ordem. Os predicados de primeira ordem

são as expressões que comumente utilizamos para dizer algo sobre os objetos denotados por diferentes termos singulares. Por exemplo, a expressão 'ser presidente' é um predicado de primeira ordem na medida em que nos permite formar frases como 'Macron é presidente'. Por outro lado, os predicados de segunda ordem são expressões com as quais formamos frases cujos sujeitos lógicos são os próprios predicados de primeira ordem. Assim, por exemplo, na frase 'A presidência é um cargo político', 'ser um cargo político' é um predicado de segunda ordem aplicado a uma versão nominalizada do predicado de primeira ordem 'ser presidente'. Da mesma forma, de acordo com Frege, a verdadeira forma lógica de um enunciado existencial 'Existe algo que é *F*' é 'o conceito *F* é instanciado', em que 'ser instanciado' é um predicado de segunda ordem aplicado ao predicado de primeira ordem 'ser *F*'. Ou seja, para Frege, o quantificador existencial não é mais do que uma versão mascarada do predicado de segunda ordem 'ser instanciável'.

A ideia de que quantificadores são tipos especiais de expressão denotativa aparece em diferentes lugares da literatura lógica. Em primeiro lugar, a concepção substitucional dos quantificadores (KRIPKE, 1976) está associada a uma perspectiva denotativa sobre o papel semântico dessas expressões linguísticas. Informalmente falando, em uma concepção substitucional dos quantificadores, um enunciado existencial $\exists x\, Fx$ e um enunciado universal $\forall x\, Fx$ teriam, respectivamente, o mesmo conteúdo de uma disjunção $Fa \vee Fb \vee Fc \vee$ e de uma conjunção $Fa \wedge Fb \wedge Fc \wedge \ldots$ infini-tárias, em que a, b, $c\ldots$ são termos singulares para cada elemento da estrutura de interpretação alvo. Ou seja, desse ponto de vista, os quantificadores seriam expressões que denotariam genericamente todo o universo de discurso.

Também Fine (1983) nos sugere uma concepção dos quantificadores como denotadores genéricos, por sua vez sem recurso a elementos infinitários. Nessa obra, Fine sugere uma visão segundo a qual, em uma sentença universal $\forall x\, Fx$, o sujeito lógico é um objeto genérico obtido a partir de um princípio de abstração das propriedades comuns a todos os elementos do universo de discurso.

Não é fácil determinar quem está certo no debate entre concepções predicativas e denotativas da semântica dos quantificadores. Note que ambas as posições são fortemente carregadas do ponto de vista teórico na medida em que igualmente dependem da assumpção de um hiato entre a forma gramatical 'superficial' dos enunciados e a sua estrutura lógica 'profunda'. Além disso, ambas as posições, nas sua diferentes vertentes, podem depender da assumpção de algum tipo de pluralismo ontológico, seja via propriedades de segunda ordem, seja pela introdução de objetos genéricos. Apesar disso, argumento aqui que, assumindo-se uma concepção denotativa dos quantificadores é possível estabelecer uma via de generalização do problema de Frege e de sua solução em termos da distinção sentido-referência a um conjunto maior de casos de falha da onisciência lógica. A disponibilidade de uma via de unificação de casos de ignorância lógica pode contar como uma razão a favor da concepção denotativa da semântica dos quantificadores.

A ignorância sobre a verdade de identidades da forma $a=b$ se revela na falha de substituição de idênticos em contextos intensionais. Por exemplo, alguém poderia saber que a Hesperus é um planeta e ao mesmo tempo acreditar que a Phosphorus não é um planeta. Um caso semelhante de falha de substituição ocorre em contextos quantificacionais. Consideremos o seguinte exemplo (de certa forma, esse é um caso bastante idealizado, mas serve como protótipo de cenários mais realistas). A fórmula

(*) $\neg(\forall x\, \exists y\, (\varphi x \wedge \neg \varphi y))$

é uma verdade lógica. O tablô semântico abaixo nos oferece uma demonstração de sua validade:

(1) $\forall x\, \exists y\, (\varphi x \wedge \neg \varphi y)$	
(2) $\exists y\, (\varphi a \wedge \neg \varphi y)$	(Instanciação universal em 1)
(3) $\varphi a \wedge \neg \varphi b$	(Instanciação existencial em 2)
(4) $\exists y\, (\varphi b \wedge \neg \varphi y)$	(Instanciação universal em 1)
(5) $\varphi b \wedge \neg \varphi c$	(Instanciação existencial em 4)

(6) $\neg \varphi b$ (Eliminação da conjunção em 3)

(7) φb (Eliminação da conjunção em 5)

X

Dito informalmente, essa demonstração explora o fato de que o quantificador existencial ocorrendo na fórmula $\forall x \exists y (\varphi x \wedge \neg \varphi y)$ oferece uma 'testemunha' que, se instanciada no quantificador universal, gera uma contradição. Logo, o indivíduo que ignora a validade da fórmula (*) ignora que o domínio do quantificador existencial está incluído no domínio do quantificador universal.[2]

No problema de Frege original, a ignorância sobre a verdade de uma identidade é ignorância sobre a correferencialidade de dois termos singulares. Similarmente, no caso quantificacional acima apresentado, a ignorância sobre a validade de (*) é ignorância sobre a co-extensionalidade dos quantificadores que ocorrem na fórmula. Portanto, pode-se definir uma versão generalizada do problema de Frege do seguinte modo:

> *Problema de Frege generalizado*: Há casos em que um agente racional ignora a validade lógica de uma fórmula porque desconhece a co-extensionalidade dos quantificadores que nela ocorrem.

O caso de ignorância sobre a validade de (*) é um exemplo (idealizado) entre outros possíveis do problema de Frege generalizado. É uma questão ainda em aberto determinar um *teorema de caracterização* do conjunto de casos de ignorância lógica que podem ser capturados pela definição generalizada do problema de Frege. Este artigo passa ao largo dessas questões mais técnicas (que devem ser investigadas em um trabalho futuro). Ao invés disso, no restante deste trabalho, me dedico a apresentar como é possível adaptar a distinção sentido-referência para o exame da semântica dos quantificadores e da versão generalizada do problema de Frege.

[2] Apesar da idealidade do exemplo (*), ele é um protótipo de casos realistas de ignorância lógica. A construção de tablôs fechados para π_2-fórmulas não é um simples procedimento de aplicação mecânica de regras sintáticas, mas envolve frequentemente engenhosidade, algo que pode faltar ao agente racional.

2. Sentido e bidimensionalismo semântico

Apesar do seu apelo, o conceito de sentido tal como apresentado por Frege não é suficientemente claro. Como dito acima, Frege entende por sentido um modo de apresentação da referência, um dispositivo semântico que cumpre o papel de identificar o objeto do discurso em um dado contexto de uso da linguagem. No entanto, isso ainda é insuficiente: apenas com base nessa explicação, não ficamos sabendo que tipo de dispositivo semântico é esse e como ele é capaz de cumprir tal função. O *bidimensionalismo semântico* promete nos oferecer um melhor entendimento do sentido das expressões denotativas.

O marco histórico inicial do bidimensionalismo semântico é o trabalho de Kaplan (1989) sobre a semântica de expressões indexicais. Expressões como 'eu', 'agora', 'aqui' possuem um duplo aspecto. Por um lado, o conteúdo dessas expressões apresenta um componente estático: em qualquer contexto de uso da palavra 'eu', essa expressão sempre denota o proferidor do ato linguístico. Por outro lado, o conteúdo de expressões indexicais apresenta igualmente um elemento dinâmico: na boca de diferentes falantes, 'eu' denota pessoas diferentes. Para explicar essa dupla faceta das expressões indexicais, Kaplan lança mão das noções de *caráter* e *conteúdo*.

O conteúdo de uma expressão é o seu valor semântico propriamente dito. Por exemplo, se Macron diz 'eu sou o presidente da França', a expressão 'eu' ocorrendo nessa frase denota, nesse contexto de uso, Macron. Logo, o conteúdo de 'eu', nesse contexto de uso da expressão, é Macron. O caráter, por sua vez, é uma função que, em cada contexto de uso, atribui à expressão dada um conteúdo semântico. Assim, o caráter da expressão 'eu' é tal que, quando essa expressão é utilizada por Macron, a função atribui o conteúdo Macron a 'eu', quando a expressão é utilizada por Barack Obama, a função atribui o conteúdo Obama e assim por diante. Logo, os conceitos de caráter e conteúdo capturam, respectivamente, os elementos estático e dinâmico da semântica de expressões indexicais.

O sucesso da análise do significado dos indexicais em termos da distinção caráter-conteúdo sugeriu, em seguida, aplicações mais generalizantes

desse par conceitual à semântica da linguagem natural. Inspirados pelo argumento da *Terra Gêmea* de Putnam (1973), filósofos externalistas passaram a sustentar que o conteúdo semântico de uma sentença qualquer pode variar entre contextos de uso da linguagem. Contudo, o externalismo semântico se associa imediatamente a um risco de ceticismo semântico: se uma mesma sentença pode ter conteúdos diferentes em distintos contextos de uso, como é possível a comunicação intercontextual? Motivado por essa questão, Stalnaker (1978) propôs uma generalização da semântica bidimensional para o corpo global da linguagem.

Para Stalnaker, toda sentença está associada a uma função caráter que, em cada contexto de uso da linguagem, lhe atribui uma proposição, a qual, por sua vez, em termos modais, pode ser descrita como uma função que leva de mundos (contextos de uso da linguagem) a valores de verdade. Assim, por exemplo, o caráter da frase 'O copo está cheio d'água' é uma função tal que, na Terra (i.e., no contexto atual de uso da linguagem) atribui à frase um conteúdo que é verdadeiro na Terra e falso na Terra Gêmea, e na Terra Gêmea atribui ao enunciado um conteúdo verdadeiro na Terra Gêmea e falso na Terra. Esse comportamento do caráter de 'O copo está cheio d'água' pode ser visualizado a partir da seguinte tabela:

	Terra	Terra Gêmea
Terra	V	F
Terra Gêmea	F	V

Tabela 1: Terra e Terra Gêmea representam diferentes contextos de uso da linguagem. As linhas da tabela representam em termos vero-condicionais os conteúdos de 'O copo está cheio d'água' em cada contexto linguístico considerado. As colunas representam o modo como cada contexto linguístico avalia cada conteúdo atribuível ao enunciado.

O que é então veiculado na comunicação intercontextual de 'O copo está cheio d'água', de acordo com Stalnaker, é a *proposição diagonal* do enunciado, i.e., o conteúdo proposicional gerado pela função ponto fixo da Tabela

1 (representado pela linha diagonal da tabela). Ou seja, para esse autor, em uma comunicação intercontextual, o que se veicula é somente um núcleo comum minimal das diferentes interpretações de um enunciado, o que, no mais das vezes, não é mais do que o seu valor de verdade atual em cada contexto linguístico. Assim por exemplo, em uma comunicação entre um terráqueo e seu *doppelgänger* da Terra Gêmea, a proposição transmitida pela asserção de 'O copo está cheio d'água' (i.e., a proposição diagonal desse enunciado) apresenta um conjunto de condições de verdade tal que tanto na Terra quanto na Terra Gêmea ela é verdadeira.

Um aparente problema da proposta de Stalnaker é o seu caráter formalista, i.e., faltam elementos a essa teoria semântica para barrar cenários em que, em algum contexto de uso linguístico, o enunciado 'O copo está cheio d'água' poderia receber interpretações selvagens como 'O copo está cheio de água tônica' (para mais detalhes, cf. SCHROETER & BIGELOW, 2009). Diante disso, autores como Jackson (2004) reforçam a leitura indexical da noção de caráter. Desse modo, em defesa do bidimensionalismo de Stalnaker, se poderia morder a bala sustentando a razoabilidade de tais cenários semânticos. Dessa perspectiva, o caráter das expressões de nossa linguagem apresentaria uma indexicalidade implícita. Por exemplo, a expressão 'água' denotaria *aquilo* que enche os rios e mares e serve para matar a sede. O que de fato cumpre essa função em cada contexto de uso da linguagem poderia variar bastante livremente.

Cabe notar que a ideia de que a nossa linguagem está repleta de elementos indexicais se associa à tese de que nossa competência semântica está acompanhada de um conhecimento imperfeito da referência das expressões denotativas. Embora sejamos plenamente capazes de utilizar a palavra 'água', nós não sabemos o que essa palavra denota em cada contexto de uso da linguagem. Nossa competência semântica não depende desse conhecimento. De fato, pode inclusive ser o caso que nós não sejamos capazes de rastrear a referência de um termo mesmo no contexto atual de uso da linguagem (esse é o caso histórico dos usuários do conceito de água antes do estabelecimento de

sua fórmula química).[3] No que segue, argumento que algo similar se passa com nossa competência no uso de quantificadores.

Antes de voltarmos ao tema dos quantificadores e ao problema de Frege generalizado, note que, concebendo-se a noção fregeana de sentido em termos do conceito bidimensionalista de caráter, podemos dar a seguinte caracterização modal do problema de Frege original e de sua solução fregeana. Seja @ o atual contexto de uso da linguagem em que os nomes 'Hesperus' e 'Phosphorus' correferem o planeta Vênus. A identidade 'Hesperus=Phosphorus' em sua interpretação atual é verdadeira em @ e, dado que os nomes próprios são designadores rígidos, é também verdadeira, nessa mesma interpretação, em todos os outros mundos possíveis. Logo, 'Hesperus=Phosphorus' em sua interpretação atual é logicamente verdadeira. No entanto, em alguma interpretação contrafactual, 'Hesperus' e 'Phosphorus' denotam objetos celestes distintos em algum mundo possível. Ou seja, sob essa interpretação deviante, 'Hesperus=Phosphorus' não é logicamente verdadeira. Portanto, o fenômeno da ignorância sobre a validade da equação 'Hesperus=Phosphorus' se deve ao fato de que a situação epistêmica de um agente racional pode ser incapaz de determinar qual é a interpretação atual do enunciado.

Essa caracterização modal do problema de Frege e de sua solução fregeana apresenta ao menos duas propriedades importantes. Em primeiro lugar, ela preserva a logicalidade do problema, i.e., essa formalização mostra claramente que o problema de Frege é um caso particular do problema da onisciência lógica. Em segundo lugar, ela mostra que, concebida em termos modais, a solução fregeana do problema da informatividade das identidades da forma a=b não exige a consideração de mundos classicamente impossíveis. Ou seja, o recurso à semântica bidimensional nos mantém no terreno da lógica modal clássica. Esse é um aspecto importante ao qual retornaremos mais adiante.

[3] Note, contudo, que essa observação é compatível com a tese defendida por alguns bidimensionalistas (e.g., JACKSON, 1994) de que é possível oferecer uma forma de análise *a priori* do conteúdo semântico de nossas expressões linguísticas.

3. Uma solução fregeana para o problema de Frege generalizado

Se assumirmos que os quantificadores são tipos especiais de denotadores genéricos, tal como pressuposto pela concepção substitucional da quantificação, por exemplo, então é plausível considerar que essas expressões linguísticas também apresentam algum tipo de indexicalidade. No que segue, argumento que nosso uso competente de quantificadores depende da atenção a caráteres que selecionam, em cada contexto de uso da linguagem, um domínio de quantificação admitindo, contudo, certa variabilidade nas características dos domínios selecionados.

Em termos modelo-teóricos, o domínio de um quantificador nunca é absoluto, mas antes está sempre relativizado a um universo de discurso *dado*. Nesse sentido, o quantificador universal denota genericamente todos os indivíduos dados, e o quantificador existencial denota genericamente alguns dos indivíduos dados. Naturalmente, o universo de discurso dado não está fixado de antemão, mas pode variar entre contextos de uso da linguagem. Se, passando do lado de fora de uma sala de aula, vemos o professor escrever no quadro-negro a fórmula '$\forall x\ (0 < x)$', não sabemos ainda qual é o domínio do quantificador universal que nela ocorre. Ele varia entre os números naturais apenas ou considera inclusive os racionais positivos? Para decidir essa questão é preciso olhar para o contexto de uso do enunciado, i.e., precisamos saber qual é o tema da aula.

De todo modo, para entender o significado de um quantificador não é necessário conhecer a configuração extensional do domínio de quantificação atualmente dado. É possível compreender o sentido da fórmula '$\forall x\ (0 < x)$' mesmo sem saber qual é o domínio de quantificação de $\forall$. Logo, é plausível sustentar que a expressão 'dado' nessa explicação informal do significado dos quantificadores envolve uma indexicalidade implícita: o que está dado pode variar entre contextos linguísticos, para conhecer a configuração extensional do domínio de quantificação atualmente dado é preciso olhar para o contexto de uso do quantificador, mas a nossa competência semântica no uso dessa expressão não requer esse conhecimento.

No entanto, se os quantificadores enquanto expressões denotativas indexicais têm caráter, quais são as condições mínimas que uma tal função impõe a conjuntos para que esses possam cumprir o papel de domínios de quantificação? Com base nos tablôs semânticos (em especial, nas regras sintáticas de instanciação universal e existencial), é plausível afirmar ao menos o seguinte:

- *Caráter do quantificador universal*: o domínio dado inclui ao menos todas as 'testemunhas' previamente consideradas.
- *Caráter do quantificador existencial*: o domínio dado inclui ao menos uma 'testemunha' nova, não considerada previamente.

Assim caracterizados, os caráteres dos quantificadores universal e existencial toleram contextos linguísticos em que algum enunciado da forma $\forall x \exists y (\varphi x \wedge \neg \varphi y)$ é verdadeiro: na medida em que as 'testemunhas' novas oferecidas pelo quantificador existencial não foram consideradas anteriormente ao nosso uso do quantificador universal, o caráter de $\forall$ não exige que elas estejam contidas no domínio por ele selecionado. Sob essa caracterização, os caráteres dos quantificadores toleram, portanto, contextos nos quais a fórmula (*) não tem sua validade lógica reconhecida. Tais contextos linguísticos podem ser modelizados através da *semântica urna* (RANTALA, 1979), um fragmento expressivo da lógica clássica que enfraquece as relações de dependência entre quantificadores ocorrendo em uma fórmula. Nesse sistema formal, é possível construir modelos impossíveis (não-clássicos), que satisfazem a fórmula $\forall x \exists y (\varphi x \wedge \neg \varphi y)$ ao não incluírem o domínio do quantificador existencial no domínio do quantificador universal. Para mais detalhes cf. Mendonça (No prelo).

Como vimos no final da seção anterior, com base no bidimensionalismo semântico pode-se oferecer a seguinte solução para o problema de Frege original: um agente racional pode não saber que $a=b$ porque a sua situação epistêmica tolera contextos linguísticos contrafactuais em que os caráteres de a e b apontam para referências distintas. Similarmente, pode-se dar a seguinte

solução bidimensional ao problema de Frege generalizado: um agente racional pode não saber que uma certa fórmula quantificada é logicamente válida porque a sua situação epistêmica tolera contextos linguísticos contrafactuais que não reconhecem a co-extensionalidade de dois ou mais quantificadores ocorrendo na fórmula. Nesses contextos linguísticos, os caráteres dos quantificadores em questão selecionam domínios não co-extensionais, ao contrário do que acontece no contexto linguístico atual. Consequentemente, em tais contextos contrafactuais a fórmula em questão é vista como expressando uma contingência e não uma verdade lógica.

Note que ao passarmos da solução bidimensional do problema de Frege original à solução bidimensional da sua versão generalizada, alguns dos recursos conceituais utilizados se sofisticaram. Em particular, enquanto na solução do problema original, os contextos contrafactuais de uso da linguagem considerados são ainda cenários clássicos do ponto de vista lógico, o mesmo não se dá quando passamos à versão generalizada do problema. Contextos contrafactuais que não reconhecem a co-extensionalidade de domínios de quantificação são necessariamente cenários impossíveis do ponto de vista da lógica clássica, i.e., a modelização formal desses cenários requer inevitavelmente a adoção de algum sistema de lógica não-clássica. Assim, por exemplo, um possível contexto em que os nomes 'Hesperus' e 'Phosphorus' não referem ao mesmo corpo celeste ainda é um cenário que respeita todas as leis da lógica clássica. Portanto, a consideração de um tal cenário não exige uma revisão da lógica. No entanto, um contexto que não reconhece que o domínio do quantificador existencial ocorrendo na fórmula $\forall x \exists y (\varphi x \wedge \neg \varphi y)$ está incluído do domínio do quantificador universal é necessariamente não-clássico.

Deve-se notar que a revisão da lógica exigida pela solução bidimensional do problema de Frege generalizado tem uma razão puramente epistemológica. É necessário revisar a lógica para que se possa melhor caracterizar a situação epistêmica dos indivíduos acerca dos fatos lógicos. No entanto, os cenários contrafactuais que podem ser assim modelizados são apenas epistemicamente possíveis, i.e., eles são tolerados pela situação epistêmica de um indivíduo que ignora que uma certa fórmula expressa uma verdade lógica, mas

eles não são de fato possíveis. Logo, de um ponto de vista ontológico, nós ainda estamos mergulhados na lógica clássica, embora reconheçamos que o estudo da relação epistemológica que os indivíduos mantêm com ela por vezes nos exige a adoção de recursos conceituais não-clássicos.

Conclusão

No presente artigo, vimos como, a partir da adoção de uma concepção denotativa dos quantificadores, é possível caracterizar uma versão generalizada do problema de Frege. Enquanto o problema de Frege original trata de casos de ignorância sobre a correferencialidade de dois termos singulares, a versão generalizada do problema por sua vez trata de casos de ignorância sobre a co-extensionalidade de dois ou mais quantificadores ocorrendo em uma fórmula. Além disso, a partir de uma análise bidimensionalista da distinção sentido-referência, foi possível sugerir uma solução fregeana para o problema generalizado.

Esse resultado é bastante relevante pois indica como se pode unificar casos do problema da onisciência lógica à primeira vista distintos entre si. No entanto, uma questão permanece: será possível fazer mais? Será possível oferecer uma generalização ainda mais profunda do problema de Frege? Me parece que sim. Por exemplo, a definibilidade infinitária dos quantificadores existencial e universal em termos dos conectivos disjunção e conjunção talvez indique uma via de generalização do problema de Frege ao caso da lógica proposicional. No entanto, uma análise mais detalhada dessa questão ainda depende de investigação futura.

Referências

FINE, K. “A defence of arbitrary objects”. *Proceedings of the Aristotelian Society, Supplementary Volumes*, v. 57, 1983, p. 55-89.

FREGE, G. *Os Fundamentos da Aritmética*. São Paulo: Abril Cultural, 1983.

_____. “Sobre o sentido e a referência”. In: *Lógica e Filosofia da Linguagem.* São Paulo: Edusp, 2009, p. 129-158.

FREIRE, R. A. “First-order logic and first-order functions”. *Logica Universalis*, v. 9, n. 3, 2015, p. 281-329.

HENDRICKS, V.; SYMONS, J. “Epistemic logic”. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab, Stanford University, 2015.

JACKSON, F. “Armchair Metaphysics”. In: MICHAEL, M.; O’LEARY-HAWTHORNE, J. (Eds.) *Meaning in Mind*. Dordrecht: Kluwer Academic Publishers, 1994, p. 23–42.

_____. “Why we need A-intensions?” *Philosophical Studies*, v. 118, n. 1-2, 2004, p. 257-277.

KAPLAN, D. “Demonstratives”. In: ALMOG, J. et al. (Eds.). *Themes from Kaplan*. London: Oxford University Press, 1989, p. 481-564.

KRIPKE, S. “Is There a Problem About Substitutional Quantification?” In: EVANS, G.; MCDOWELL, J. (Eds.), *Truth and Meaning*. London: Oxford University Press, 1976, p. 324-419.

MENDONÇA, B. R. “New semantics for urn logics: taming the enduring scandal of deduction”. *No prelo.*

PUTNAM, H. “Meaning and reference”. *The Journal of Philosophy*, v. 70, n. 19, 1973, p. 699-711.

RANTALA, V. “Urn models: a new kind of non-standard model for first-order logic”. In SAARINEN, E. (Ed.). *Game-Theoretical Semantics*. Amsterdam: Springer, 1979, p. 347–366.

SCHROETER, L.; BIGELOW, J. “Jackson’s classical model of meaning”. In RAVENSCROFT, I. (Ed.). *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*. Oxford: Oxford University Press, 2009.

STALNAKER, R. “Assertion”. *Syntax and Semantics*, v. 9, 1978, p. 315–332.

SOBRE ELIMINAR A IDENTIDADE DA LÓGICA

Hudson dos Anjos Benevides[1]

Resumo

Neste artigo argumentarei que não se pode afirmar sem qualificação que a identidade pode ser removida da lógica como proposto por Wehmeier (2004). Isto é feito ao apontar o fato básico de que a identidade não é definível na linguagem da lógica de primeira ordem. Como sugerido por Béziau (2013), em lógica de primeira ordem temos no máximo a Identidade de Leibniz—uma relação de congruência e não a identidade numérica. Portanto, qualquer tentativa de dispensar a identidade em termos de equivalência lógica entre linguagens de primeira ordem não pode ser bem sucedida. Além disso, apresentamos uma desvantagem metafísica ao mudar para a lógica defendida por Wehmeier (2004).

Palavras-chave: Identidade. Dispensabilidade. Aridade. Logicidade.

ABSTRACT

In this paper I will argue that one cannot claim without reservations that identity can be removed from logic as proposed by Wehmeier (2004). This is done by pointing out the very basic fact that identity is not a notion definable in the language of first order logic. As suggested by Béziau (2013), in first order logic we have at most Leibniz's Identity— a congruence relation and not numerical identity. Therefore, it is argued, any attempt to dispense identity in terms of logical equivalence between languages of first order cannot succeed. In addition, we present a metaphysical drawback in changing to the logic advocated by Wehmeier (2004).

Keywords: Identity. Dispensability. Arity. Logicality

Introdução

A despeito de ser comum apresentar a identidade como parte da lógica, seu estatuto lógico é controverso. Um modo para mostrar que uma noção

[1] Doutorando em Filosofia pelo Programa de Pós-graduação Lógica e Metafísica da Universidade Federal do Rio de Janeiro. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.
Email: benevideshudson@gmail.com. ORCID: 0000-0003-4705-7509.

não é lógica é mostrar que ela não desempenha nenhum papel central em certas argumentações, que seu papel lógico é dispensável. O dado para o filósofo da lógica e qualquer um interessado nos limites entre aspectos lógicos e não lógicos é o seguinte: como pode a identidade ser dispensável ao mesmo tempo que, por diferentes critérios, ela tenha sido considerada uma noção lógica? Geralmente critérios de logicidade são esboçadas com vistas a um conjunto de termos tradicionalmente considerados lógicos. Qual força poderia ter certo critério se o conjunto selecionado pudesse envolver alguma noção para qual existam razões suficientes para dispensá-la? Embora a prática comum seja demarcar o campo das noções lógicas olhando para grupos de noções tradicionalmente consideradas lógicas, poucos filósofos têm reconhecido a necessidade de uma análise conceitual individual que busque de fato entender porque tradicionalmente certas noções têm sido assim consideradas. Há, assim, a necessidade de investigação acerca das consequências das disputas sobre a dispensabilidade da identidade para o quadro geral de demarcação entre noções lógicas—um empreendimento que tradicionalmente têm sido visto como central para a compreensão sobre a natureza da lógica. Ao compreender os sentidos específicos em que a identidade é logicamente dispensável avançamos, portanto, ao esclarecimento sobre o as características fundamentais de uma noção lógica.

O presente artigo tem como objetivo apresentar uma avaliação crítica da tese de eliminabilidade da identidade como apresentada em Wehmeier (2004). Em linhas gerais, a posição apresentada é a de que podemos nos livrar de todos os usos do símbolo de identidade e seu apelo à alegada noção problemática de identidade numérica—a relação que um objeto possui consigo mesmo e nada mais—ao adotarmos um novo arcabouço formal: a lógica de predicados wittgenstainiana (LPW). Apresento a seguir duas razões básicas para questionar essa posição, a saber, que o aparato formal utilizado para se livrar da identidade se livra, no máximo, daquilo do que podemos chamar de Identidade de Leibniz-Béziau (2013).

A estrutura argumentativa do presente texto é a seguinte. Começamos com os esclarecimentos básicos sobre o que é a identidade e seu papel lógico

(seção 1). Em seguida, mostramos como podemos eliminar as ocorrências do símbolo de identidade das formalizações que supostamente envolvem a identidade (seção 2). Após ter ilustrado este ponto, passamos a discussão sobre por que não é possível argumentar que com a remoção das ocorrências do símbolo de identidade obtemos a eliminabilidade da identidade (seção 3). Por fim (seção 4), sugerimos que um problema surge ao apoiarmos a explicação do estatuto de verdade necessária de afirmações de identidade em termos de um operador de correferência.

1. O papel lógico da identidade

Para que a ideia defendida neste trabalho se torne mais clara, nesta seção apresentamos o que vem a ser o papel lógico da identidade dado que isto será útil na próxima seção onde será mostrado como esse papel pode ser dispensado.

A identidade é um predicado introduzido na linguagem da Lógica de Primeira Ordem (LPO) por meio do símbolo "=". Diferente de operadores como conjunção ($\wedge$) ou disjunção ($\vee$) que conectam sentenças simples para formar sentenças complexas, o símbolo identidade forma uma sentença a partir de dois termos, como por exemplo variáveis ou constantes individuais. Assim, podemos tomar duas variáveis x e y e produzirmos a sentença $x = y$, querendo com isto dizer que os valores daquelas variáveis são idênticos, da mesma forma que na linguagem corrente usamos dois nomes como 'Túlio' e 'Cícero' ou 'Hesperus' e 'Phosphorus' para formar sentenças como 'Túlio é Cícero' e 'Hesperus é Phosphorus'. Assim, por noção lógica de identidade entendemos aqui o operador que permite expressar as afirmações do tipo mencionado.

Como notado, a identidade é um predicado, mas se trata de um predicado especial. Predicados, em geral, podem ter diferentes extensões de acordo com diferentes interpretações em LPO. No entanto, a identidade mantém seu significado constante. Por isso, pensamos nesse predicado como uma constante lógica. Como veremos adiante, parte da razão para ser concebida enquanto noção lógica é o papel que desempenha tanto na formalização de sentenças

quanto na regimentação de certas expressões linguísticas. Qual seria, portanto, o significado da identidade?

Costumeiramente, o sinal de identidade denota o que podemos chamar de *identidade numérica*—a relação binária que todo objeto tem consigo mesmo e nada mais (NOONAN; CURTIS, 2014). Na lógica de primeira ordem com identidade ($LPO^{=}$), ao especificar as condições de verdade para as sentenças, temos que uma sentença do tipo $t_1 = t_2$ é verdadeira, se e somente se o valor semântico de t_1 é o mesmo que o de t_2, onde t_1 e t_2 são termos da linguagem.[2] Do ponto de vista modelo-teórico, a identidade numérica pode sempre ser definida sob qualquer conjunto de objetos como a diagonal do produto cartesiano daquele conjunto de objetos—o conjunto de pares ordenados da forma x , x. Por isso, se a identidade é uma constante lógica, ela deveria, à princípio, sempre ter como extensão a diagonal do domínio do discurso.

A identidade entra como parte do vocabulário porque sem ela não seríamos capazes de formalizar certas afirmações da linguagem natural. Em outras palavras, os conectivos proposicionais junto com quantificadores, sozinhos, são insuficientes para expressar as afirmações que envolvem a identidade numérica. Por exemplo, a afirmação

Por que se importar com o significado da identidade? Por que alguém deveria ser compelido a adicionar a identidade no conjunto do vocabulário lógico da LPO? Existem duas respostas viáveis para essas questões. Por um lado, podemos enfatizar o *papel* que a identidade desempenha em formalizações do raciocínio ordinário e regimentações de noções da linguagem natural que de outro modo não poderiam ser feitas. Por outro lado, podemos buscar uma resposta mais profunda, querendo saber em virtude do que a identidade é uma noção lógica; ou seja, buscamos uma justificativa para o estatuto lógico da identidade. Como exemplo de um critério desse tipo (e estritamente formal), podemos citar a proposta de Tarski (1986). Segundo Tarski (1986, p. 149) um noção é lógica apenas quando ela é "invariante sob todas as possíveis

[2] Os termos podem ser pensados como análogos formais das expressões que em linguagem natural usamos para referir a objetos. Para fins de simplificação, "termos" dizem respeito a variáveis, para os quais usaremos letras minúsculas como x,y,z , ou constantes individuais, para as quais usaremos as letras a,b,c . Além disso, letras maiúsculas, como F, G H, são usadas para predicados.

transformações um-a-um do mundo em si mesmo", o que ficou conhecido como o critério da *invariância sob permutação*. Seguindo a proposta de Tarski (1986), há apenas um grupo seleto de relações binárias que contam como lógicas, entre elas a identidade e a diversidade/diferença. Como a tese da (in)dispensabilidade da identidade discutida adiante diz respeito apenas ao uso da identidade para formalização do raciocínio ordinário, iremos nos concentrar aqui no *papel* lógico da identidade ao invés de seu *status* lógico[3].

(1) Hesperus é Phosphorus

onde a ocorrência da identidade não é explícita, como quando dizemos 2 + 2 = 4, é formalizada utilizando o símbolo da identidade . Essa formalização é, na forma mais simples,

(2) $h = p$

onde h é uma constante individual que denota Hesperus e p a constante individual que denota Phosphorus. Outras afirmações que supostamente envolvem a identidade são:

(3) Hesperus existe

que pode ser formalizada como:

(4) $\exists x(x = h)$

tendo h o mesmo referente que antes. Igualmente envolvendo a identidade temos afirmações que envolvem descrições definidas, por exemplo,

(5) O indivíduo que escreveu *Cinzas do Norte* é amazonense

[3] Para um explicação e articulação da posição tarskiana veja (BELLOTTI, 2003). Para uma discussão geral e aprofundada do tópico sobre constantes lógicas veja (GÓMEZ-TORRENTE, 2002)

cuja a formalização, agora padrão,

(6) $\exists x(E\,xc \wedge \forall\, y(Eyc \rightarrow y = x) \wedge Ax)$

deriva do tratamento russelliano das descrições definidas. Por fim, temos o caso de expressões que indicam cardinalidade, como as do tipo "no mínimo *n*" em

(7) No mínimo duas coisas existem

que podemos formalizar como

(8) $\exists x\, \exists y(\neg\, x = y)$

e para expressões do tipo "no máximo *n*", como em

(9) No máximo dois livros foram vendidos

onde para dizer que a quantidade de livros não ultrapassa dois, parte da formalização deve incluir

(10) $\forall\, z((Lz \wedge Cz) \rightarrow z = x \vee z = y)$

supondo que haja x e y que são livros e foram comprados.

Esses casos ilustram o que tomamos aqui como o papel lógico da identidade.[4] Abstraindo a partir destes casos, podemos ver que do ponto de vista da linguagem da LPO$^{=}$, a identidade lida com a coincidência de valores entre variáveis, ou entre valores e constantes individuais, ou ainda com casos mistos entre variáveis e constantes individuais. É importante notar que a lista não pretende ser exaustiva, muito menos que esta lista seja a razão mais relevante para a tese da indispensabilidade da identidade. Ao longo do desenvolvi-

[4] Também poderia ser incluído o papel inferencial da identidade por meio, por exemplo, de suas regras em dedução natural.

mento da lógica diferentes razões têm sido propostas para sustentar a indispensabilidade da identidade.

Dummett (1973, p. 542) sugere que Frege tenha sido o primeiro a introduzir a identidade como uma noção lógica. Em linhas gerais, a razão para tanto teria sido o fato de que qualquer definição explícita envolveria a identidade. No entanto, a identidade seria tão básica que para defini-la seria de antemão pressupô-la. [5] Além disso, Dummett (1973, p. 22) sugere outra motivação consistente com alguns pressupostos fregeanos, a saber, a indispensabilidade da identidade para expressar as assim chamadas *condições quantificacionais*—expressões que indicam algo sobre a cardinalidade dos objetos do domínio, como nos exemplos (7)-(10). Contudo, não seria possível encontrar outros modos de lidar com os cenários que envolvem a identidade? Certamente, se esse cenário for possível ele põe em cheque uma respeitável tradição em lógica.

Certamente a identidade é uma noção importante. Hodges (2001, p. 66) enfatiza que ela é necessária para todas as aplicações matemáticas da lógica. Não por acaso a tese da indispensabilidade remonta aos trabalhos pioneiros de Frege e Russell onde a preocupação com a linguagem natural é paralela à preocupação com os fundamentos da matemática. No entanto, se indispensabilidade da identidade significa *indispensabilidade para formalizar argumentos e regimentar expressões da linguagem natural*, então aparentemente possuímos uma justificativa frágil para tanto.

2. A fragilidade da tese da indispensabilidade

Existe uma técnica no tratamento de variáveis que permite a eliminação do símbolo de identidade na LPO. Por conta disso, parece plausível que possamos argumentar, em princípio, que a identidade é eliminável da lógica. Adiante serão apresentadas razões para duvidar dessa afirmação, por isso o assunto desta seção é a eliminabilidade formal da identidade.

[5] Para uma discussão sobre a tese da indefinibilidade da identidade cf. (SAVELLOS, 1990), (MCGINN, 2000) e (BUENO, 2014)

Contemporaneamente, Wehmeier (2012) tem proposto que podemos viver sem a identidade ao endossarmos a assim chamada Lógica de Predicados Wittgensteiniana—um sistema formal com poder expressivo similar ao da tradicional LPO mas que dispensa os usos da identidade. A lógica wittgensteiniana, por brevidade LPW, é baseada em grande parte em um diagnóstico feito por Hintikka (1956). Segundo a análise de Hintikka (1956) parte do que nos constrange ao uso da identidade em LPO é fruto da convenção sobre o tratamento de variáveis feito na linguagem. No contexto da LPO, é permissível que variáveis diferentes coincidam em valor. Ou seja, duas variáveis no escopo de um quantificador não necessariamente designam dois objetos diferentes. Um exemplo pode facilitar a compreensão. A afirmação

(11) Todo mundo ama alguém

poderia ser formalizada em $\text{LPO}^{=}$ ao menos de duas maneiras, a saber,

(12) $\forall x \exists y (Axy)$
(13) $\forall x \exists y (Axy \wedge \neg x = y)$

cuja a sutil diferença é a seguinte. Por conta da peculiaridade da LPO mencionada acima quanto ao tratamento das variáveis, a formalização expressa em (12) permite um cenário em que a sentença é verdadeira caso o domínio do discurso seja composto apenas por narcisistas, isto é, na circunstância em que os indivíduos não amam ninguém além de si mesmos. De forma simples, o mundo dos narcisistas torna verdadeira (12). Por outro lado, quem declara (11) pode ter outra intenção, talvez querendo excluir exatamente o cenário dos narcisistas. Assim, para que sentença seja verdadeira em um mundo de não narcisistas, precisamos da identidade (e da negação). Para esse caso, (13) parece mais apropriada. Na presente situação, a identidade entra em questão para determinar que os valores das duas variáveis não irão coincidir. A peculiaridade aqui discutida é o que Hintikka (1956, p. 226) chamou de uma

interpretação inclusiva de variáveis. [6] Dado que o valor das variáveis pode coincidir, somos impelidos a recorrer à identidade para os casos onde essa coincidência é indesejável. No entanto, existem casos onde essa coincidência é desejável, por exemplo, para expressar unicidade como em (6), e mais uma vez é necessário o uso da identidade. Hintikka (1956) mostrou que ao mudar o tratamento das variáveis para uma *interpretação exclusiva*, não permitindo que diferentes variáveis coincidam em valor, podemos ter uma linguagem com o mesmo poder expressivo que LPO, mas que dispensa os usos da identidade.

É claro, lidar com a coincidência ou dissidência entre os valores de variáveis é apenas um dos usos para a identidade e a abordagem de Hintikka (1956) possuía a limitação de tratar apenas da linguagem de primeira ordem sem constantes individuais. A lacuna deixada foi preenchida recentemente pela proposta de Wehmeier (2004, 2012) de uma lógica wittgensteiniana, referida anteriormente. Esse desenvolvimento reforça o diagnóstico Hintikka (1956), que vê na convenção do tratamento das variáveis o motivo técnico que tornaria os filósofos ou lógicos cativos da identidade.

Como exatamente a mudança no tratamento de variáveis torna a tese da dispensa a identidade em sua plenitude? Segundo Wehmeier (2012, p. 765), o princípio básico da LPW é o de que sempre que em uma sentença temos uma variável ligada a um quantificador o valor das variáveis livres ou de constantes individuais também presentes naquela sentença passam a ser excluídos do escopo daquele quantificador. Ou seja, se x está ligada ao quantificador universal ($\forall$) e a variável z não, então o possível valor para z passa a ser excluído do escopo do quantificador ($\forall$). Por exemplo, a sentença (12) acima é verdadeira na LPW quando para qualquer indivíduo a no domínio existe algum b —que *não* é o próprio a — e a ama b . Ou seja, na LPW , (12) não pode ser verdadeira no mundo dos narcisistas, onde indivíduos amam apenas a si mesmos. Portanto, a identidade não precisa mais entrar em cena para evitar

[6] Dado que o termo "interpretação" possui um significado técnico preciso, no decorrer do presente texto usarei o termo "tratamento" que convenientemente evita ambiguidade conceitual.

a coincidência das variáveis—isso já é subproduto do tratamento exclusivo das variáveis.

Quais são os efeitos práticos em se adotar a LPW e sua interpretação exclusiva de variáveis? Basicamente, todas as ocorrências do símbolo de identidade passam a ser dispensadas. Para (1) Hesperus é Phosphorus e (3) Hesperus existe, cuja formalização envolve o uso de constantes individuais, a LPW usa um operador de correferência simbolizado com a barra tripla "≡". Com esse operador, as antigas formalizações (2) e (4) tornam-se, respectivamente (14) $h \equiv p$ e (15) $\exists x\, x \equiv h$. Para as afirmações (5) e (9) obtemos respectivamente

$$(16) \exists x \left(Axc \wedge Mx \right) \wedge \neg \exists y \exists z \left(Ayc \wedge My \wedge Azc \wedge Mz \right)$$
$$(17) \neg \exists z \exists w \exists k \left(Lz \wedge Cz \wedge Lw \wedge Cw \wedge Lk \wedge Ck \right)$$

Wehmeier (2012) não fornece um tratamento direto para casos como "(7) No mínimo duas coisas existem". Com a substituição de "=" por " ≡", obtemos $\exists x \exists y \neg\, x \equiv y$. No entanto, visto na LPW diferentes variáveis nunca coincidem em valor, esta formalização passa a ser trivial. Como nada do que será dito adiante depende deste ponto, ignoramos a questão. É importante notar que o que Wehmeier (2012) propõe não é apenas uma paráfrase como ilustrado nesta seção. O resultado técnico mais geral é apresentado Wehmeier (2004), onde demonstra-se que há uma tradução de LPO na LPW, e que portanto, ao eliminar as ocorrências do símbolo de identidade não se perde poder expressivo.

Se as linguagens sob discussão são equivalentes, salvo a diferença de vocabulário, que tipos de razões alguém pode ter ao eliminar a identidade da lógica? Afinal, porque viver sem a identidade? A motivação central de Wehmeier (2012) é a de que a identidade é uma noção problemática. Como o autor enfatiza, a identidade é tradicionalmente pensada com a relação binária que uma coisa tem consigo mesma. O problema é: como pode a identidade ser uma relação binária se ela nunca pode relacionar dois objetos? Essa questão ecoa uma intuição já articulada por Wittgenstein no *Tractatus* quando afirma

"dizer de duas coisas que elas são idênticas, é sem sentido, e dizer de uma coisa que ela é idêntica a si própria, não diz absolutamente nada (5.5303)" Assim, a fundamental vantagem em excluir a identidade do vocabulário lógico seria eliminar uma noção problemática do seio da lógica. Outra vantagem em se eliminar a identidade em favor da LPW, estaria no fato de a LPW se adequar melhor ao usos da linguagem natural, seja pelo fato de que existem casos de afirmações da linguagem natural que quando formalizadas com a identidade permitem derivar certos absurdos (WEHMEIER 2012, p. 766) ou pelo fato de que as afirmações de identidade do tipo (1) Hesperus é Phosphorus seriam apenas alegações de correferência seguindo a evidência de certos desenvolvimentos na linguística (WEHMEIER 2012, p. 767). Por questões de escopo, a discussão sobre a plausibilidade dessas vantagens será postergada.

A seguir, apresentamos razões duvidar da ideia de que com a traduti-bilidade da $LPO^{=}$ na LPW estamos eliminando, de fato, a identidade da lógica.

3. Qualificando a dispensabilidade da identidade

Será mesmo que a identidade é dispensável? Na presente seção uma resposta negativa é dada a essa questão. Mais precisamente, o que será argumentado é que a identidade não é dispensável nos moldes apresentados anteriormente (seção 2) porque o arcabouço formal utilizado por Wehmeier (2012) não é o lugar próprio da identidade. Seguindo Béziau (2013), podemos afirmar que o que é dispensado com a LPW é simplesmente a identidade de Leibniz.

Hintikka (1956) Wehmeier (2012) discordam sobre quais conclusões tirar dos resultados acerca da dispensabilidade da identidade. Hintikka (1956, p. 228) não vê qualquer vantagem na interpretação exclusiva da variáveis que tem por efeito a dispensabilidade da identidade, pois ela nos dá um nova forma de "dizer exatamente as mesmas coisas que podíamos dizer antes". Wehmeier (2012, p. 764), por outro lado, sugere que a vantagem da dispensabilidade é que ela torna o sinal de igualdade "e com ele o apelo à

identidade numérica como seu valor semântico, supérfluo". O que se segue pode ser entendido como uma defesa do ponto de vista de Hintikka.

Pretendo argumentar que Hintikka está certo e que não há qualquer vantagem do tipo alegada por Wehmeier (2012). O motivo é simples: a identidade não é uma noção de primeira ordem—temos a intenção de capturá-la em primeira ordem, mas na maioria dos casos isso não é possível. Além disso, a introdução do operador "≡" para dar conta de casos onde relacionamos constantes individuais introduz problemas que não estão presentes quando tomamos o papel da identidade como básico. A seguir, discutimos em detalhe a ideia de que a identidade não é uma noção de primeira ordem.

Como bem sabido existem alguns princípios lógicos que regem "=", a saber,

$$(18)\ \forall x x = x$$
$$(19)\ \forall x \forall y \left(x = y \rightarrow Fx \leftrightarrow Fy\right)$$

composta pelo princípio da *reflexividade da identidade* (18) e lei a *indiscernibilidade dos idênticos* (19), por vezes conjuntamente chamadas de Leis de Leibniz. [7] No entanto, esses princípios não podem ser considerados axiomas da identidade, pois não podemos encontrar um conjunto de axiomas sobre "=" cujo os modelos serão sempre a identidade numérica ((HODGES, 2001, p. 64), Béziau (2013, p. 2)). A consequência disso é que apenas com (18) e (19) não se pode garantir que "=" sempre terá como extensão a diagonal do domínio do discurso, que é um modo de conceber a identidade numérica do ponto de vista modelo-teórico. Portanto, a ideia de que junto da eliminação do símbolo evitamos ao mesmo tempo o apelo à identidade numérica como seu valor semântico deve ser devidamente qualificada. Que "=" vá sempre referir à identidade numérica, é algo dado quase por convenção. Em outras palavras,

[7] A nomenclatura pode variar dependendo da literatura. (HODGES, 2001, p. 65) considera ambas como leis de Leibniz. Béziau (2013) usa algo como (20) acima , que pressupõe a identidade dos indiscerníveis, como Lei de Leibniz, ao passo que (NOONAN; CURTIS, 2014) referem-se apenas ao princípio de indiscernibilidade dos idênticos como Lei de Leibniz.

do ponto vista formal, o que é garantido é que "=" apela quase sempre a identidade de Leibniz e não à identidade numérica.

Béziau (2013), por exemplo, usa o fato de a identidade numérica não ser axiomatizável em primeira ordem contra a tese de Quine (1986) de que um dos motivos para a identidade ser uma noção lógica seria o fato de que ela pode ser eliminável. A proposta de eliminação proposta por Quine é ligeiramente diferente. Segundo Quine (1986), com um conjunto finito de predicados, podemos fornecer um substituto para a identidade. Por exemplo, se linguagem possui apenas dois predicados, podemos simular o comportamento da identidade usando a equivalência:

$$(20) x = y \leftrightarrow \left(Fx \leftrightarrow Fy \wedge Gx \leftrightarrow Gy \right)$$

e para Quine isso mostra que a identidade está mais próxima da lógica, já que seu papel pode, em última instância ser simulado com os conectivos lógicos já presentes na linguagem da LPO. Segundo Béziau (2013, p. 3), se fizermos a distinção sutil entre a identidade numérica e a identidade de Leibniz, Quine não poderia concluir que está eliminando *a* identidade numérica, mas a identidade de Leibniz.

Parece razoável considerar que a mesma posição pode ser mantida quanto a vantagem da eliminabilidade da identidade sugerida por Wehmeier (2012). Ora, o máximo que pode ser alegado é que com a eliminabilidade do símbolo se torna supérfluo o apelo a identidade de Leibniz. Ou seja, essa eliminabilidade não é forte o suficiente para mostrar porque deveríamos eliminar a identidade de uma vez por todas do vocabulário lógico, ou mais que isso acatar a LPW como o novo padrão.

Alguém pode objetar que a dificuldade em captar a identidade em uma linguagem de primeira ordem mostra que ela seria uma noção problemática. No entanto, essa objeção parece perder de vista a real dificuldade, a saber, que o que está em jogo não é a identidade, mas a capacidade de poder usar o aparato formal da LPO em capturar a identidade. Em outras palavras, a dificuldade em fornecer axiomas para a identidade em primeira ordem é uma

limitação da lógica e não da identidade. Como Quine (1986) acertadamente notou, a identidade é uma noção universal. Qualquer que seja o domínio do discurso, se estamos falando de pessoas, cadeiras ou números, a identidade não conhece preferências. Um modo de entender essa afirmação, como sugere Béziau (2013, p. 3), é pensar na identidade nos termos modelo-teóricos apresentados no início deste texto: a diagonal do produto cartesiano do domínio do discurso. É claro que podem haver outras razões para achar que seja vantajoso eliminar a identidade, ao se mostrar, por exemplo, que identidade é uma noção problemática. No entanto, esse movimento muda a dialética da discussão. Parece razoável pensar que, embora intimamente relacionadas, possamos tratar separadamente a questão se a identidade é eliminável da lógica das razões para se duvidar que exista uma relação não problemática de identidade. Além disso, não só a tradução da LPO na LPW de primeira ordem não possui a força que deveria para eliminar a identidade, mas também, se fosse bem sucedida parece ter a consequência de forçar certos compromissos ontológicos desnecessários e que não se fazem presentes com o uso da identidade.

4. Sobre a necessidade de afirmações de identidade

Wehmeier (2012, p. 774) reconhece a seguinte dificuldade em sua abordagem. Uma das contribuições de Kripke (1980) é a de que afirmações como (1) Hesperus é Phosphorus são afirmações necessárias, isto é, elas são verdadeiras em todos os mundos possíveis. Ao analisar afirmações de identidade entre nomes próprios em termos do operador "≡", na proposta de Wehmeier (2012), o caráter necessário daquelas afirmações aparentemente se perde. Nesta seção pretendo argumentar que a tentativa de recuperar o caráter necessário de afirmações de identidade na LPW imputa compromissos desnecessários aos que declaram afirmações como (1). É importante notar que as dificuldades apresentadas a seguir são de cunho metafísico. Elas não necessariamente dizem respeito a uma falha na argumentação proposta por Wehmeier, mas sim apontam consequências metafísicas indesejadas ao adotar o arcabouço teórico por ele defendido.

Por que exatamente o caráter necessário de afirmações como (1) Hesperus é Phosphorus se perde? Em linhas gerais, outros planetas poderiam ter sido nomeados de "Hesperus" e "Phosphorus", ou seja, esses nomes poderiam falhar em correferir. Sendo assim, afirmações com a forma $h \equiv p$não são sempre verdadeiras, e portanto, não poderiam ser vistas como verdades necessárias. Para fugir desse problema, sem fazer uso da identidade, Wehmeier (2012, p. 774) sugere que a intuição básica que induz o pensamento que afirmações como (1) são verdadeira é a de que "é impossível para um objeto ser dois objetos". Em sua perspectiva, afirmações do tipo "a é b" são parafraseadas como "a e b são um objeto", e "tudo o que isto envolve é um discurso elementar sobre cardinalidade que pode ser facilmente acomodado na presença de um módico de lógica de segunda-ordem" (WEHMEIER 2012, p. 774). De acordo com isso, ao afirmar (1) Hesperus é Phosphorus alguém estaria no fundo dizendo que (21) "Hesperus é Phosphorus são um objeto". E, através das lentes da LPW, com o auxílio da lógica de segunda ordem, (21) seria formalizada como:

$$(22) \exists F\left(\left(Fh \wedge Fp\right) \wedge \neg \exists y \exists z\left(Fy \wedge Fz\right)\right)$$

que deve ser lida como dizendo que "podemos formar uma coleção à qual pertencem h e p, mas que não possui mais de um membro" (WEHMEIER 2012, p. 774). Essa análise seria correta pois de acordo com Wehmeier (2012), (22) é uma consequência de um princípio lógico em LPW; o de que todo objeto tem a propriedade de singularidade (*singleton property*), isso é expresso como:

$$(23) \forall x \exists F\left(Fx \wedge \neg \exists y \exists z\left(Fy \wedge Fz\right)\right)$$

outra sentença de ordem superior. Qual seria as consequências indesejadas da proposta em seguir usando a LPW para explicar o caráter necessário de afirmações como (1)? Como já sinalizado, a primeira diz respeito ao compromisso ontológico dessas abordagens. Outra, proposta por Trueman (2014) é a de

que toda essa análise parece minar o próprio projeto de se livrar da identidade.

O resultado de concordar com a ideia de que (1) Hesperus é Phosphorus deve ser analisada como (21) "Hesperus e Phosphorus são um objeto", e por consequência que (2) $h \equiv p$é equivalente a (22) $\exists F((Fh \wedge Fp) \wedge \neg \exists y \exists z(Fy \wedge Fz))$é de que ao afirmar (1) estamos nos comprometendo não apenas com o objeto referido por "Hesperus" e "Phosphorus", mas também com a existência de uma classe. Para aqueles que, como Quine (1963), desejam evitar compromissos ontológicos desnecessários, essa análise pode parecer equivocada: (1) é sobre Hesperus ser numericamente idêntico a Phosphorus, isto é, sobre os objetos designados por aqueles nomes e não sobre a *existência de uma classe* unitária de objetos. Qual seria o compromisso ontológico indesejado aqui? Classes são tipos de entidades abstratas, ou seja, são tipos de entidades que não possuem localização espaço-temporal e causalmente inertes. Filósofos céticos quanto a existência de tais entidades, seriam forçados pelo novo arcabouço a se comprometer com aquilo que pretendem evitar. Obviamente, isso pode levantar a suspeita de que são esses filósofos que devem fornecer um tratamento alternativo à eliminação da identidade. O parágrafo seguinte fornece um esboço de como isso poderia ser feito.

Como observam Humberstone e Townsend (1994, pg. 253), "equivalência lógica não significa equivalência conceitual", e ao que parece (22) é uma extrapolação conceitual. Humberstone e Townsend (1994) propõem que façamos a distinção entre o *conceito de identidade* e a *relação de identidade*. Por conceito identidade os autores concebem algo como "a é (numericamente) idêntico a b". Em linhas gerais, Humberstone e Townsend (1994, pg. 260) defendem que podemos entender que no exercício do conceito de identidade somos constrangidos pelas regras de inferência que envolvem o predicado "=", "ninguém que compreenda o sentido de = poderia falhar em ver a correção de afirmações da forma $t = t$, ou duvidar da propriedade de inferir, por exemplo, Fb a partir de Fa juntamente com $a = b$". Nada disso nos conduz ao comprometimento com a existência de uma classe qualquer que seja. Ou seja, nada

impede que nossas intuições sobre o estatuto de afirmações de identidade derivam da identidade conceitual ao invés da identidade numérica. E ao parafrasear "a é b" como "a e b são um objeto" só estamos tornando explícito que "a é (numericamente) idêntico a b". Portanto, um arcabouço conceitual em que identidade é uma noção básica, evita compromissos indesejados.

Por fim, migrar para lógica de segunda ordem para explicar um fato básico como a necessidade de afirmações de identidade é desvantajoso para quem pretende dispensá-la. Como observado por Trueman (2014, p. 166), a identidade pode ser definida e, mais que isso, axiomatizável, em lógica de segunda ordem.[8] O modo tradicional de fazê-lo é por meio de algo como:

$$(24) \forall x \forall y (x = y \leftrightarrow \forall F (Fx \leftrightarrow Fy))$$

Sendo assim, como pode a identidade ser dispensável da lógica se é possível logicamente definí-la? Wehmeier (2012, 774, nota 21) possui uma resposta para isso: basta traduzir (24) na LPW de segunda ordem, o resultado que obtemos é o seguinte:

$$(25) \forall x (Ixx \leftrightarrow (\forall F (Fx \leftrightarrow Fx) \wedge \forall x \forall y (Ixy \leftrightarrow \forall F (Fx \leftrightarrow Fy))))$$

onde $I(x, x)$e Ixy, pretendem capturar o predicado de identidade, ou seja, devem ser interpretados como "x é idêntico a x" e "x é idêntico a y". No entanto, como Wehmeier (2012, p. 773) sugere, na LPW de segunda ordem Ixxe Ixy seriam respectivamente notações alternativas "supérfluas" para tautologia (Taut(x)) e contradição (Con(x, y)). Essa abordagem, como Trueman (2014, p. 167) acertadamente observa é dialeticamente irrelevante para o propósito de eliminar a identidade. O fato de que (24) se torna (25) na LPW de segunda ordem "nos impressionará só se já estivermos convencidos a usar [LPW]". Mais uma vez, tradutibilidade entre LPO e LPW não possui força suficiente para dispensar a identidade.

[8] Cf. SHAPIRO, S. **Foundations Without Foundationalism**: A Case for Second-Order Logic. Oxford University Press, 1991.

Existem outras maneiras de motivar a necessidade de dispensar a identidade. Entre elas está a de tentar mostrar que o caráter binário da relação de identidade é problemático: como pode a identidade ser binária se ela nunca relaciona dois objetos? Essa é outra das estratégias que Wehmeier (2012) usa na tentativa de dispensar a identidade. Analisar a efetividade das razões para tanto foge do escopo do presente trabalho. A pergunta que pretendemos como fio condutor é a de se há como eliminar a identidade da lógica, o que plausivelmente não parece ser o caso.

Considerações Finais

Com as considerações precedentes parece razoável que não é fácil eliminar a identidade da lógica, se isso significar mudar de uma lógica para outra que elimina todas as ocorrências do símbolo de identidade. Fica claro que é preciso cuidado com certas distinções conceituais e com a significância dada ao aparato formal para as discussões em filosofia da lógica, em especial, para a credibilidade da tese de que a identidade é eliminável da lógica.

Na seção 1 buscamos apresentar para quais tipos de afirmações a identidade é indispensável e que qualquer pretensa maneira de eliminar a identidade deveria dar conta. Em seguida, na seção 2, apresentamos uma maneira de se livrar da identidade, a saber, utilizando a lógica de predicados wittgenstainiana proposta por Wehmeier (2012). As seções restantes, apresentam a contribuição do presente artigo. Na seção 3 argumenta-se que a identidade não é eliminável da lógica dado que o aparato formal utilizado for de primeira ordem. Na seção 4 argumenta-se que o estatuto modal de afirmações de identidade não podem ser recuperados sem o custo de imputar compromissos ontológicos desnecessários.

Se correto, isso indica que devemos avaliar cuidadosamente o quanto se pode filosoficamente extrapolar a partir de certo aparato formal nas discussões em lógica. É importante notar que o raciocínio empregado ao longo do texto é basicamente em termos modelo-teóricos. Fica em aberto a significância da semântica demonstrativo-teórica (*proof-theoretics*) para o papel lógico da identidade. Regras de inferência seriam suficientes para salvaguardar o pa-

pel privilegiado da identidade? Haveria algum substituto inferencial para a identidade? Estas questões demandam investigação que está para além dos modestos propósitos deste artigo.

Referências

BELLOTTI, Luca. "Tarski on logical notions". *Synthese*, v. 135, n. 3, p. 401-413, 2003.

BÉZIAU, Jean-Yves. "Quine on Identity". *Principia* , v. 7, n. 1-2, p. 1–15, 2013.

BUENO, Otavio. "Why Identity Is Fundamental". *American Philosophical Quarterly*, v. 51,n. 4, p. 325–332, 2014.

DUMMETT, Michael. *Frege: Philosophy of Language*. New York: Harper & Row, 1973.

GÓMEZ-TORRENTE, Mario. "The problem of logical constants". *Bulletin of Symbolic Logic*, v.8, no. 1, pp. 1-37, 2002.

HINTIKKA, Jaakko. "Identity, Variables, and Impredicative Definitions". *Journal of Symbolic Logic*, v. 21, n. 3, p. 225–245, 1956.

HODGES, Wilfrid. "Elementary Predicate Logic". In: GABBAY, Dov; GUENTHNER, Franz. (Ed.). *Handbook of Philosophical Logic*. London: Springer. pp. 1–129, 2001.

MCGINN, Colin. *Logical Properties: Identity, Existence, Predication, Necessity, Truth*. USA: Oxford University Press, 2000.

NOONAN, Harold.; CURTIS, Ben. "Identity". In: ZALTA, Edward N. (Ed.). *The Stanford*

Encyclopedia of Philosophy. Summer 2014. [S.l.]: Stanford, 2014. Disponível em: <https://plato.stanford.edu/archives/sum2018/entries/identity/>

QUINE, Willard v. O. "On what there is". In: QUINE, Willard v. O. *From a Logical Point of View*. New York: Harper Torchbooks, 1963. p. 1–19

_____. *Philosophy of Logic* . 2nd. [S.l.]: Harvard University Press, 1986.

SAVELLOS, Elias. "On Defining Identity". *Notre Dame Journal of Formal Logic* , Duke University Press, v. 31, n. 3, p. 476–484, 1990.

TARSKI, Alfred. “What are logical notions?” In: CORCORAN, John. (Ed.). History and Philosophy of Logic 7: 143–154, 1986.

TRUEMAN, Robert. “Eliminating identity: a reply to Wehmeier”. *Australasian Journal of Philosophy*, v. 92, n. 1, pp. 1-8, 2014

WEHMEIER, Kai. “How to Live Without Identity—And Why”. *Australasian Journal of Philosophy*, v. 90, n. 4, p. 761–777, 2012.

_____. “Wittgensteinian Predicate Logic”. *Notre Dame Journal of Formal Logic*, v. 45, n. 1, p. 1–11, 2004.

WITTGENSTEIN, Ludwig. *Tractatus Logico-Philosophicus and Philosophical Investigations*. Tradução: M. S Lourenço. Tratado Lógico-Filosófico e Investigações Filosóficas. Lisboa: Fundação Calouste Gulbenkian, 1987.

UMA INTRODUÇÃO AOS PENSAMENTOS SINGULARES SOBRE OBJETOS MATEMÁTICOS

Pedro Carné[1]

RESUMO

Apresento, neste artigo, uma introdução à possibilidade de pensarmos singularmente objetos matemáticos. Para isso, após caracterizar em linhas gerais tanto os pensamentos singulares, quanto o problema dos pensamentos singulares sobre objetos matemáticos, reconstruo duas teorias comprometidas com esta possibilidade, sejam elas as teorias de James Davies e Jody Azzouni.

Palavras-chave: Pensamentos Singulares. Objetos Matemáticos. Arquivos Mentais. Canais Informacionais.

ABSTRACT

This paper outlines a portrayal of our capacity of thinking mathematical objects in a singular way. In order to do that, I analyze two theories that commit themselves to the truth of this capacity, be them the theories presented by James Davies and Jody Azzouni.

Keywords: Singular Thought. Mathematical Objects. Mental Files. Information Channels.

Introdução

O objetivo deste artigo consiste em apresentar uma introdução à possibilidade de pensarmos singularmente objetos matemáticos. Para isso, propõe-se o seguinte itinerário.

Em primeiro lugar, procura-se caracterizar os pensamentos singulares através do comportamento psico-linguístico dos termos singulares. Apesar de não ser exaustiva, esta caracterização é suficiente para delinearmos os pensamentos cujas especificações encontram-se em discussão.

Na sequência, indicamos alguns argumentos de Kent Bach contrários à existência de pensamentos singulares acerca de objetos matemáticos. Com

[1] Doutor em Filosofia pela PUC-RJ e professor da UACS/UFCG.
E-mail: pedrohpcarne@gmail.com. ORCID:.0000-0002-7692-8441.

esta menção, pretende-se dar a conhecer, ainda que brevemente, um posicionamento contrário à possibilidade em discussão.

As duas seções seguintes apresentam argumentos favoráveis à existência de pensamentos singulares sobre objetos matemáticos. Nomeadamente, os argumentos desenvolvidos por James Davies e Jody Azzouni. A razão pela qual concentramo-nos nos argumentos destes autores consiste no fato de eles nos permitirem contemplar, em relação aos pensamentos singulares, o antagonismo entre platonismo e nominalismo em filosofia da matemática.

Por fim, faz-se o esboço de algumas questões e perspectivas suscitadas por esta pesquisa.

A definição dos pensamentos singulares através do comportamento psico-semântico de seus termos

A apresentação dos pensamentos singulares através do comportamento psico-linguístico dos termos singulares representa, talvez, a maneira mais simples de que dispomos para explicitar o contraste que esta espécie de pensamento ou maneira de pensar manifesta em relação aos pensamentos gerais ou descritivos.[2]

Considere-se, por exemplo, a diferença entre os pensamentos expressados pelas sentenças "Leão Tolstói morreu em 1910" e "O autor de *Guerra e Paz* morreu em 1910." Os defensores da tese singularista argumentam que a primeira sentença nos autoriza a pensar em Tolstói de maneira singular, uma vez que nele pensamos por intermédio de seu *nome próprio*, ao passo que a segunda sentença nos faz pensar em Tolstói de maneira descritiva, na medida em que lançamos mão da descrição definida "o autor de *Guerra e Paz*". Isto é, a diferença entre *nomes próprios* e *descrições definidas* auxilia-nos na elaboração da diferença entre pensamentos singulares e pensamentos descritivos, pois enquanto estas fazem referência ao escritor russo de maneira indireta, através da utilização abreviada de quantificadores e propriedades, aqueles nos permitem falar de Tolstói diretamente, de maneira determinada.

[2] Esta estratégia manifesta-se, por exemplo, em Azzouni 2009, Jeshion 2010b e Crane 2013.

O comportamento psico-linguístico dos pronomes demonstrativos também nos ajudam a elaborar semelhante distinção. Com efeito, os pronomes demonstrativos são capazes de nos mostrar ou de nos apresentar um determinado objeto sem nomeá-lo diretamente. Essa capacidade é caracterizada como *função deíctica* e constitui, de acordo com algumas análises gramaticais, o traço definidor dessa classe de pronomes (CUNHA e CINTRA, 2008, p. 342-356). Isto significa que o mero emprego de um pronome demonstrativo é suficiente para situarmo-nos em relação a um certo objeto. Assim, ao encontrarmo-nos diante de uma garrafa laranja em cima da mesa, tanto podemos pensar que há um objeto que instancia a propriedade de ser garrafa, ser laranja e encontrar-se em cima da mesa (acrescendo-se que somente um objeto instancia todas essas propriedades conjuntamente), quanto pensar que o fato de esta garrafa laranja encontrar-se em cima da mesa origina uma sentença tal como "esta garrafa laranja em cima da mesa".

De todo modo, os *pensamentos singulares* são usualmente caracterizados como *pensamentos não-descritivos*, os quais, naturalmente, contrapõem-se aos *pensamentos descritivos*. Em outras palavras, os pensamentos singulares constituem uma espécie de pensamento ou uma maneira de pensar por intermédio da qual nós nos relacionamos com um determinado objeto sem o auxílio de uma conceitualização geral ou descritiva do mesmo.

Por que não poderíamos pensar singularmente os objetos matemáticos?

É inegável que possuímos pensamentos singulares sobre os mais variados objetos do mundo, afirma Kent Bach. Se assim não os pensássemos, continua, nunca nos relacionaríamos com nenhum objeto específico em nosso pensamento: nossa visão do mundo seria inteiramente qualitativa (Cf. BACH, 2010, p. 39). Naturalmente, deve ser motivo de atenciosa investigação de que modo somos capazes de pensar singularmente um determinado objeto.

A melhor maneira de apreciarmos o problema dos pensamentos singulares, para Bach, é relacioná-lo com um problema que lhe seria análogo, seja este o problema da particularidade de um objeto na percepção. Suponhamos, por exemplo, que você esteja lendo este artigo em seu celular. Provavelmente,

você está vendo o seu celular na sua frente. Apesar disso, diz-nos Bach, diversos filósofos acreditam na possibilidade de você possuir a mesma experiência que possui agora (no momento em que olha para o seu celular) ainda que fosse outro o celular que estivesse na sua frente, ainda que fosse outro o objeto, ou ainda que nada estivesse na sua frente, como se a presença do objeto não fosse essencial para a constituição da experiência subjetiva do mesmo. A conclusão extraída destes cenários por tais filósofos é a de que "o conteúdo de todas as experiências perceptuais apenas podem ser gerais" (*ibid.*, p. 40). Efetivamente, os argumentos de Bach caminham na contramão de tais diagnósticos.

Bach afirma que, "certamente, quando percebemos algo, estamos conscientes *disso*, e não apenas de que existe uma coisa única de um certo tipo" (*idem*). Nossa experiência de um objeto concreto, continua, não seria semelhante à visão de uma sombra ou de uma pegada. Tampouco seria semelhante à visão da mancha de caneta em uma camisa. Nestes casos, nossa relação com o objeto responsável pela sombra ou pela pegada, assim como com a caneta cuja tinta manchou a camisa, só se processaria com o auxílio de descrições. Como, porém, poderíamos explicar nossa relação com um objeto independentemente de uma descrição? Isto é, o que haveria na experiência de um objeto que demandaria a presença deste objeto específico, e não a presença de quaisquer outros objetos possuidores de características idênticas? Em linhas gerais, Bach defende que é suficiente, para pensarmos singularmente um determinado objeto, estarmos conectados com alguma representação do mesmo, uma vez que tais representações funcionariam como "indexicais mentais". E afirma: "existem diferentes tipos de relações representacionais *de re*, baseadas na percepção, memória ou comunicação; porém, por mais remotas e conectadas que sejam, [estas relações] devem estar fundamentadas na percepção do objeto representado por parte de algum indivíduo" (*ibid.*, p. 61).

A simples associação do problema dos pensamentos singulares ao problema (análogo) da particularidade na percepção permite-nos assinalar um traço fundamental dos pensamentos singulares para Bach: seu caráter causal. O caráter causal dos pensamentos singulares implicaria por sua vez na tese se-

gundo a qual só pensamos desta maneira objetos cuja percepção ocorreu em algum momento de nossa existência (caso não tenhamos percebido o objeto em questão, é fundamental que o mesmo tenha sido percebido por outra pessoa). E, com o perdão da brevidade, se prosseguirmos neste raciocínio, e reconhecermos que os objetos matemáticos são objetos abstratos, concluiremos a impossibilidade de pensarmos singularmente objetos matemáticos. De fato, nas palavras de Bach (1987, p. 12), "a relação que faz de algo um objeto de pensamento *de re* é uma relação causal […]. Entidades abstratas simplesmente não podem entrar em relações causais".[3]

Este não é o espaço para investigarmos pormenorizadamente a teoria de Bach. Para nossos propósitos, é suficiente ressaltarmos que, em sua concepção, um traço fundamental dos pensamentos singulares é o seu caráter causal. Particularmente, o apelo à percepção é suficiente para compreendermos a interdição a qualquer tratamento aos objetos abstratos. Embora Bach reconheça o fato de sua teoria ter sido elaborada com a intenção de explicar nossos pensamentos singulares acerca de objetos concretos (ainda que, em sua opinião, tais objetos não constituam a totalidade dos objetos que poderíamos pensar singularmente), ela se tornou uma espécie de paradigma da impossibilidade de pensarmos singularmente os objetos abstratos. Naturalmente, isso não impediu que muitos esforços fossem feitos para contradizer seus argumentos.

Pensamentos singulares sobre objetos matemáticos: o poder explicativo dos arquivos mentais

Nesse sentido, o primeiro ponto de vista favorável aos pensamentos singulares sobre objetos matemáticos que gostaria de explorar encontra-se em um recente artigo publicado por James Davies. Nele, Davies argumenta que nenhuma das três principais teorias contemporâneas sobre a singularidade de nossos pensamentos — teoria da *acquaintance*, instrumentalismo semântico e

[3] Posteriormente, Bach justifica sua decisão de não elaborar em maiores detalhes, ainda que negativamente, nossa impossibilidade de pensar singularmente objetos abstratos: "não discuti a referência a entidades abstratas porque acredito que os problemas reais não se encontram na relação de referência, mas no estatuto ontológico e epistemológico destas entidades" (1987, p. 261).

cognitivismo semântico — é capaz de explicar satisfatoriamente a razão pela qual estaríamos aptos a pensar os objetos matemáticos de maneira singular.[4] Em sua opinião, se estamos de fato interessados em compreender de que forma poderíamos assim pensar tais objetos, deveríamos examinar minuciosamente alguns aspectos da teoria dos arquivos mentais, especificamente as condições para desenvolver e manter um arquivo mental sobre um objeto percebido no passado, uma vez que tais condições auxiliariam na compreensão do caso dos objetos matemáticos.

As teorizações sobre a referência e o pensamento singulares deveriam respeitar, o máximo possível, nossas intuições mais comuns e habituais sobre o comportamento psico-semântico dos termos com os quais desenvolvemos tais pensamentos, de acordo com Davies. Assim, na medida em que nossa intuição tende a considerar o termo com o qual designamos um objeto matemático como um termo singular, ao invés de considerá-lo como uma descrição definida (seja disfarçada, seja abreviada), Davies afirma (2019, p. 4) que "deveríamos procurar uma explicação que permita [que os termos numéricos] funcionem como termos singulares — pelo menos até que seja mostrado que não existe uma explicação plausível". Não é bem essa a direção de seu artigo, no entanto. Ao invés de elaborar o funcionamento dos termos numéricos como termos singulares, o artigo de Davies, por um lado, defende a existência de pensamentos singulares sobre objetos matemáticos baseado em nossas intuições sobre o comportamento psico-semântico dos termos numéricos, e, por outro lado, exibe de que forma tais pensamentos seriam possíveis com o auxílio da teoria dos arquivos mentais.

Uma vez que Davies concentra seus maiores esforços na especificação da maneira através da qual a teoria dos arquivos mentais viabilizaria os pensa-

[4] Ao subscrever semelhante julgamento — de que a teoria da *acquaintance*, o instrumentalismo e o cognitivismo semânticos seriam as *principais* teorias contemporâneas sobre a natureza de nossos pensamentos singulares —, Davies tão somente repercute a configuração teórica esboçada por Robin Jeshion em vários de seus trabalhos. Desta forma, se quisermos entender os fundamentos desta configuração, devemos buscar um maior esclarecimento nos trabalhos de Jeshion. Contudo, para não nos alongarmos nas questões preliminares, remetemos o leitor para o artigo de Jeshion intitulado "Pensamento Singular: Acquaintance, Instrumentalismo Semântico e Cognitivismo" (2010c), no qual, por um lado, são entrelaçadas suas críticas ao caráter extremamente restritivo do conceito de *acquaintance* e à liberalidade excessiva do instrumentalismo semântico, e, por outro lado, sua teoria da autoridade cognitiva é apresentada de maneira mais detalhada.

mentos singulares sobre objetos matemáticos, vejamos alguns aspectos desta teoria.

Os arquivos mentais são costumeiramente definidos como uma espécie de depósito mental de informações que julgamos serem sobre um determinado objeto (ainda que as condições sob as quais um arquivo mental efetivamente se refira a um objeto varie de teoria para teoria), e sua relação com nossos pensamentos seria, por assim dizer, constitutiva. Isto é, pensamos em um objeto (seja de maneira singular, seja de maneira descritiva) se e somente se, no processo de pensamento, ativamos ou utilizamos um arquivo mental singular que se lhe refira. O fato de existirem tanto arquivos mentais *singulares*, quanto arquivos mentais *descritivos* torna perceptível que Davies simplesmente traduz em termos de arquivos mentais a diferença entre pensamentos singulares e descritivos, sem avançar seja na especificação de ambas as maneiras de pensar, seja na especificação da natureza destes "arquivos mentais", seja na especificação do significado desta "ativação" dos arquivos mentais no processo de pensamento. Com efeito, após traduzir a diferença entre pensamentos singulares e descritivos em termos de arquivos mentais, a estratégia argumentativa adotada por Davies apresenta-nos uma subdivisão de cada um destes ramos. Isto é, os arquivos mentais singulares podem ser *demonstrativos* ou *estáveis*, ao passo que os arquivos mentais descritivos podem ser *puros* ou *holísticos*.

É forçoso traçar o seguinte parêntese: na complexa engenharia conceitual posta em movimento para dar conta do caso dos pensamentos singulares sobre objetos matemáticos, devemos inserir três teses subscritas (não necessariamente fundamentadas) por Davies: (i) que um arquivo mental "vazio" é desprovido de referência; (ii) que o mero surgimento de um arquivo mental em nossa cognição implica a possibilidade de o ativarmos ou utilizarmos conscientemente; e (iii) que a identidade de um arquivo mental é transparente para o indivíduo que o utiliza (*sabemos* se um arquivo mental é igual ou diferente em relação a outro).[5]

[5] A fundamentação da transparência de um arquivo mental, no texto de Davies, decorreria da "chancela" de Recanati a um comentário introdutório de Laura Schroeter no artigo "A ilusão da transparência", de acordo com o qual "a maioria dos filósofos da mente aceita a tese […] de que possuiríamos um acesso transparente ao conteúdo de nossos pensamentos: da-

Pois bem, a diferença entre os dois tipos de arquivos mentais singulares, para Davies, é genética, ou seja, manifesta-se na origem da informação por eles armazenada. Como os arquivos mentais baseiam-se em certas relações que estabelecemos com os objetos no ambiente, sejam elas relações epistemicamente recompensadoras (por permitir-nos obter informações diretamente a partir do objeto), os arquivos mentais demonstrativos seriam aqueles que existem durante a relação perceptual com o objeto, isto é, eles existem enquanto se mantém a relação que possibilita o fluxo de informação do objeto para o arquivo. Já os arquivos mentais estáveis, naturalmente, não necessitariam da manutenção de nenhuma relação perceptual com o objeto para se manterem em nossa cognição.

Desta distinção entre arquivos mentais demonstrativos e estáveis, Davies conclui o que nos parece uma obviedade: quando pensamos singularmente um objeto por intermédio de um arquivo mental estável, isso significa que não necessariamente o estamos percebendo no momento presente. Isto é, como o arquivo mental é estável, podemos não nos encontrar em uma relação perceptual com o objeto de nosso pensamento — o caso mais emblemático dessa situação seriam os objetos percebidos no passado. Esta possibilidade é explicada através de uma "conversão" sofrida por um arquivo mental demonstrativo por intermédio da qual ele se metamorfosearia em um arquivo mental estável. Como nem todos os arquivos mentais demonstrativos convertem-se em arquivos mentais estáveis, Davies procura especificar o papel desempenhado pela "retenção de informação" que possibilitaria esta conversão, por acre-

do que sejamos minimamente racionais, não podemos simplesmente confundir um conteúdo conceitual por outro" (SCHROETER, 2007, p. 597). Desta afirmação, Davies conclui sumariamente que, "se isto está correto, então, se parece ao indivíduo que ele está reativando o mesmo arquivo mental uma segunda vez, é isso que ele está fazendo. Consequentemente, a identidade do arquivo mental é transparente ao indivíduo" (DAVIES, 2019, p. 21). No entanto, deve-se registrar que o comentário de Schroeter, na íntegra, é o seguinte: "O senso comum nos diz que normalmente possuímos um acesso direto e confiável aos conteúdos de nossos pensamentos conscientes. Mas a maioria dos filósofos da mente aceita a tese bem mais forte de que possuiríamos um acesso transparente ao conteúdo de nossos pensamentos: dado que sejamos minimamente racionais, não podemos simplesmente confundir um conteúdo conceitual por outro" (SCHROETER, 2007, p. 597). Para ela, aliás, nada tão forte quanto a "transparência" é defendido pela psicologia contemporânea. Mesmo Recanati não admite em termos absolutos a transparência de um arquivo mental para um indivíduo, haja vista as análises apresentadas em "Transparência e os seus limites" e em "Factividade, transparência e Fraca co-referência *de jure*" (respectivamente RECANATI, 2012, p. 133-143; e RECANATI, 2016, p. 14-31). Sem apelar para uma concepção irrestrita da "transparência", todavia, a análise de Davies se enfraqueceria substancialmente.

ditar que aí encontrará a chave para desembaraçar o caso dos pensamentos singulares sobre objetos matemáticos. Sua especificação se traduzirá em duas teses.

A primeira refere-se ao fato de a retenção de informação ser uma condição necessária e suficiente para a conversão de um arquivo mental demonstrativo em um arquivo mental estável. Sua necessidade decorre de uma tese derivada da própria definição dos arquivos mentais, a de que um arquivo mental "vazio" não possuiria nenhuma referência. Se um arquivo mental vazio não possui referência, conclui Davies, a falha na retenção de informação sobre um objeto resulta na ausência de um arquivo mental que se lhe refira. E a suficiência? Para Davies (*ibid.*, p. 14), "reter informação que assumimos ser sobre um objeto individual *é apenas* reter um arquivo mental. Assim, se um sujeito retém informação sobre um objeto na ausência de uma ligação perceptual, então ele está mantendo um arquivo mental estável para este objeto". Ou seja, Davies está nos dizendo que a conversão de um arquivo mental demonstrativo em um arquivo mental estável decorre simplesmente do fato de que, por um lado, retemos informação sobre um objeto qualquer, e, por outro lado, se retemos informação sobre ele é porque há um objeto cujas informações retidas justificaria a existência do arquivo mental.

A segunda tese refere-se a uma possibilidade mais ousada. Davies pensa que não apenas os arquivos mentais demonstrativos podem ser convertidos em arquivos mentais estáveis, mas que haveriam arquivos mentais *descritivos* que também poderiam ser convertidos em arquivos mentais estáveis, tornando possível, a uma descrição, auxiliar no desenvolvimento de pensamentos singulares. De acordo com Davies (*ibid.*, p. 16), o resultado desta conversão de arquivos mentais descritivos seria "um arquivo mental estável sobre o objeto que satisfaz a descrição no momento da conversão — assim como o referente de um arquivo mental estável derivado de uma percepção é o referente de um arquivo mental demonstrativo convertido no momento da conversão". Pois bem, como essa conversão de um arquivo mental descritivo em um arquivo mental estável ocorreria?

Para Davies, é fundamental que, em primeiro lugar, concebamos uma descrição definida satisfeita por apenas um objeto. A partir desta descrição, e, talvez, de algumas outras crenças que possuímos acerca deste objeto, inferimos algumas informações sobre o objeto descrito, informações que, por alguma razão, retemos em nossa mente. Davies, então, afirma:

> Dado que o sujeito está retendo um estoque de informações sobre o objeto, sua cognição converte a descrição definida em um arquivo mental estável que se refere ao objeto que satisfaz semelhante descrição. E assim, dado que a posse de um arquivo mental estável que se refere a um objeto é necessária para tornar alguém capaz de possuir pensamentos singulares sobre o mesmo, o sujeito está agora a meio caminho de possuir pensamentos singulares sobre o objeto que satisfaz a descrição (*ibid.*, p. 17).

O cenário desenhado por Davies é animador. Todavia, há nele algumas lacunas teóricas. Afinal, Davies investigava uma subdivisão dos arquivos mentais em demonstrativos e estáveis e explicava-nos de que maneira ocorreria a conversão de um (o demonstrativo) em outro (o estável) quando lançou um novo problema para sua proposta, seja ele uma nova conversão, isto é, a conversão de um arquivo mental descritivo em um arquivo mental (singular) estável. E isso, devemos acrescentar, sem especificar de qual arquivo mental descritivo estamos falando.

Como anteriormente indicado, os arquivos mentais descritivos podem ser puros ou holísticos. Um arquivo mental puramente descritivo exibiria especificamente uma descrição definida sobre o objeto em questão, de modo que qualquer mudança na configuração desta descrição alteraria substancialmente o arquivo mental. Em outras palavras, qualquer mudança na descrição acarretaria o fim de um arquivo mental puramente descritivo bem como o início de outro. Por outro lado, um arquivo mental holisticamente descritivo apresentaria em sua constituição um conjunto de descrições para a identificação de um objeto, de modo que ele sobreviveria a uma mudança no estilo "navio de Teseu" — no sentido em que este arquivo sobreviveria a modificações paulatinas em suas descrições constituintes. Dito de outra forma, um arquivo mental holisticamente descritivo não sobreviveria a uma mudança radical de todas as suas descrições de uma só vez, mas permaneceria estruturalmente

inalterado em função de pequenas substituições de seus elementos.[6] Mas qual a diferença entre um arquivo mental singular e um arquivo mental descritivo, afinal?

Davies distingue um arquivo mental singular de um arquivo mental descritivo em função dos critérios que regeriam sua permanência em nossa cognição. Mais especificamente, a distinção entre ambas as espécies de arquivos é traçada por intermédio de uma análise do processo de modificação nas informações neles armazenadas. Para Davies, enquanto os arquivos mentais singulares seriam fortemente expressivos, em função da origem de sua informação (uma relação direta com o objeto), os arquivos mentais descritivos seriam menos expressivos, pois sua conexão com o objeto dependeria de uma única informação indireta: a relação de satisfação da descrição a partir da qual o arquivo se originou. Um arquivo mental descritivo não sobreviveria, em princípio, a uma mudança nesta descrição (ou permaneceria em um constante processo de empalidecimento, no caso dos arquivos mentais holisticamente descritivos).

Pois bem, apesar das muitas lacunas, o cenário no qual o estudo de Davies se movimenta é constituído pelas seguintes teses: (i) dependemos de arquivos mentais singulares (demonstrativos ou estáveis) para pensar de maneira singular um determinado objeto; (ii) o conteúdo dos arquivos mentais é transparente para o indivíduo que o utiliza (o indivíduo sabe se está utilizando um arquivo mental em dois momentos diferentes ou se está utilizando dois arquivos mentais distintos); (iii) um arquivo mental singular demonstrativo converte-se em um arquivo mental singular estável por intermédio de uma retenção de informação (cujas condições de possibilidade não foram pormenorizadas); (iv) uma descrição definida, que intuitivamente originaria um arquivo mental (puramente ou holisticamente) descritivo pode ser convertida em um arquivo mental singular estável. Com este instrumental teórico, vejamos a situação imaginada por Davies para dar sentido à sua proposta.

Imagine uma comunidade matemática para a qual a melhor descrição dos números naturais dá-se através dos axiomas Dedekind-Peano, de acordo

[6] Rachel Goodman (2016) elabora uma discussão mais detalhada desta caracterização dos arquivos mentais descritivos.

com os quais o número 0 seria o único número que não seria o sucessor de outro número, o número 1 seria o sucessor de 0, o 2 o sucessor do 1 e assim por diante. Subitamente, esta comunidade passa a acreditar que os números seriam melhor descritos pelas definições de Frege-Russell, isto é, que o número 0 seria o conjunto dos conjuntos equinuméricos ao conjunto de objetos não-idênticos a si mesmos, o número 1 seria o conjunto de todos os conjuntos equinuméricos à 0, o número 2 seria o conjunto de todos os conjuntos equinuméricos à 0 ou equinuméricos à 1 etc. Esta crença, adverte-nos Davies, não é o resultado de uma inferência feita a partir dos axiomas Dedekind-Peano: ela pode ter se originado de avanços ou desenvolvimentos da teoria dos conjuntos aliados a uma parcimônia ontológica. Até aí, *pas de probleme*. Davies então acrescenta a este cenário a seguinte informação (*ibid.*, p. 20): "após esta mudança, *parece subjetivamente* aos membros desta comunidade que, quando eles pensavam e falavam sobre os números naturais, eles realmente estavam falando e pensando sobre conjuntos de conjuntos equinuméricos a todo momento, à despeito das aparências em contrário".

Com isto em mente, Davies pede que imaginemos um indivíduo específico que pertença a esta comunidade. Isto é, um indivíduo que, em primeiro momento, admite os axiomas Dedekind-Peano, e, em um segundo momento, reconhece a verdade das definições de Frege-Russell. Por conseguinte, o arquivo mental deste indivíduo para o número 0 apresentaria, em um primeiro momento, informações como "é o único número que não é o sucessor de nenhum outro número", e, em um segundo momento, informações como "é o conjunto dos conjuntos equinuméricos a $\{y: y \neq y\}$". Davies acrescenta que, se *parece subjetivamente* a este indivíduo que seus pensamentos sempre se pautaram pelas definições Frege-Russell, isto significa que seus pensamentos, antes e depois da mudança paradigmática, partilhariam do mesmo conteúdo (deve-se recordar que, para Davies, em um sentido estrito de conteúdo, para que dois pensamentos expressem o mesmo conteúdo, seus arquivos mentais devem ser idênticos). Agora, sim, estamos diante de um sério dilema: como é possível que pareça subjetivamente a um indivíduo que seu arquivo mental permanece o mesmo, ainda que tenha sofrido semelhante transformação em

seu conteúdo? No caso, estamos falando de um arquivo mental cuja referência é um número que, em um primeiro momento, era caracterizado com o auxílio dos axiomas de Dedekind-Peano, e, em um segundo momento, passa a ser caracterizado com as definições Frege-Russell.

Davies avalia que, se a descrição definida que caracteriza um número específico originar um arquivo mental (pura ou holisticamente) descritivo, não poderia parecer subjetivamente ao indivíduo que utiliza este arquivo que, nos dois momentos diferentes, este arquivo possuiria o mesmo conteúdo. Pois, se um arquivo mental descritivo não sobrevive a uma mudança na descrição que o originou, não poderia parecer subjetivamente ao indivíduo que ele está ativando o mesmo arquivo mental em dois momentos distintos, uma vez que a transparência dos arquivos mentais exibiria o fim de um arquivo (baseado nos axiomas Dedekind-Peano) e o início de outro (baseado nas definições Frege-Russell). Deste modo, se a identidade do arquivo mental é transparente para o indivíduo, e se parece subjetivamente ao indivíduo que o conteúdo de seu arquivo mental é o mesmo antes e depois da mudança (dos axiomas Dedekind-Peano para as definições Frege-Russell), este indivíduo está ativando o mesmo arquivo mental em duas ocasiões distintas, e este arquivo mental não pode ser descritivo (ou, pelo menos, não pode ser puramente descritivo). Porém, por qual razão poderia parecer subjetivamente a um indivíduo que um arquivo mental possuiria o mesmo conteúdo antes e depois de a definição de número ter se modificado? Afinal, no exemplo de Davies, a modificação do conteúdo do arquivo é um fato e a percepção subjetiva do indivíduo ainda está por se justificar.

Aqui, Davies recorre a uma confusa validação para sua tese, pois afirma (*ibid*., p. 22) que "uma razão para pensarmos que pode [parecer subjetivamente ao sujeito que ele está utilizando o mesmo arquivo mental] é que não temos de procurar muito longe para ver filósofos e matemáticos sustentando afirmações que, em sua leitura mais óbvia, implica que esta aparência é possível". Nesta direção, não seria incorreto identificar o coração da análise de Davies com a afirmação inaugural de Azriel Levy em seu livro *Basic Set Theory* (1979, p. 3, meus grifos), segundo a qual "todos os ramos da matemática são

desenvolvidos, consciente ou *inconscientemente*, em teoria dos conjuntos". Davies interpreta esta afirmação de Levy não apenas como uma anuência, mas como uma corroboração da possibilidade de um indivíduo preservar o mesmo arquivo mental para um objeto matemático independentemente de quaisquer reconceitualizações sofrida por tal objeto.

Em suma, Davies afirma que duas teses viabilizariam a possibilidade de desenvolvermos pensamentos singulares sobre objetos matemáticos abstratos:

> Primeiro, que para possuir tais pensamentos é necessário que nós tenhamos arquivos mentais estáveis para tais objetos. Argumentei que isso não requer nada mais do que possuir pensamentos singulares baseados na memória. Segundo, para estes arquivos mentais estáveis serem dispositivos de pensamento singular requer que eles sejam capazes de sobreviver simultaneamente a uma completa mudança nas descrições que fixam sua referência; algo que é intuitivamente possível (*op. cit*., p. 23).

Apesar das lacunas teórico-conceituais do argumento de Davies, bem como de sua tácita aceitação de um "essencialismo oculto" naquilo que se refere aos objetos matemáticos, sua análise incorpora dois fortes pressupostos para defender a existência de pensamentos singulares sobre objetos matemáticos: por um lado, a distinção entre pensamentos singulares e descritivos em termos do conteúdo de pensamento; por outro lado, o fato de que os pensamentos singulares *dependem* dos objetos pensados para se sustentar.

Pensamentos direcionados para objetos sobre (inexistentes) objetos matemáticos

Não seria exagerado afirmar que Azzouni rejeita em sua análise cada um dos pressupostos assumidos por Davies. Isto é, para Azzouni, nem os pensamentos singulares serão caracterizados em termos de seu conteúdo, nem serão dependentes do objeto, e muito menos encontrarão seu fundamento na teoria dos arquivos mentais. A possibilidade de pensarmos singularmente um objeto matemático — tal como um número natural — será por ele expressada nos seguintes termos (2011, p. 60): "o mero uso de um nome-numérico em domínio público é suficiente para ativar pensamentos direcionados para obje-

tos mesmo sem um (genuíno) canal informacional". Vejamos, então, os elementos movimentados por Azzouni na elaboração desta sua proposta.

A maneira mais frutífera para estudarmos o fenômeno da singularidade em nossos pensamentos consistiria no exame atento do papel cognitivo por eles desempenhado, julga Azzouni. Isto significa reconhecer um divórcio entre a "referência" e a "atinência" de nossos pensamentos, uma vez que "referência é, por definição, uma relação com um objeto existente; [ao passo que] atinência é a mera representação de alguma coisa no pensamento, ela existindo ou não" (CRANE, 2011, p. 23). Consequentemente, para pensarmos singularmente, precisaríamos possuir, em primeiro lugar, alguma coisa *em mente*.

A perspectiva cognitiva permite a Azzouni afirmar que os pensamentos singulares não precisam ser somente sobre objetos existentes, como pretendera, por exemplo, John McDowell (que define um pensamento singular como aquele que não seria possível caso o objeto pensado não existisse) (Cf. 1998, p. 204). Em seus termos (AZZOUNI, 2011, p. 46), "quando penso "Mickey Mouse foi inventado por Walt Disney", estou pensando em algo específico (Mickey Mouse); estou pensando nele exatamente da mesma maneira que estou pensando no específico Walt Disney. E posso fazer isso apesar de meu total conhecimento de que Mickey Mouse não existe em sentido algum".

É importante ressaltar que, ao estudar os pensamentos singulares com ênfase em seu comportamento cognitivo, Azzouni afasta completamente duas teses defendidas por Davies, a de que tais pensamentos dependem do objeto e a de que ele seriam melhor caracterizados em termos de seu conteúdo. Isto é, a justificativa para o fato de nossos pensamentos singulares não serem caracterizados em termos de seu conteúdo caminha de braços dados com a razão pela qual tais pensamentos não seriam dependentes da existência do objeto. Fundamentalmente, para Azzouni, o que define um pensamento singular é a maneira como pensamos e não o conteúdo pensado. Os pensamentos singulares seriam especificados por intermédio de uma aguda observação de seus traços psíquicos, de seu caráter episódico, e não através de seu conteúdo proposicional. Em função disso, eles seriam melhor caracterizados como "pensamentos direcionados para objetos", pois, assim como podemos pensar

singularmente um objeto singular, também poderíamos pensar "singularmente" uma pluralidade de objetos. Aliás, o fato de podermos pensar pluralidades de maneira singular constrói a ponte teórica que possibilitará a rejeição dos arquivos mentais como o fundamento psíquico de tal maneira de pensar.

Um breve parêntese faz-se necessário: penso que qualquer apreciação da relação entre os arquivos mentais e os pensamentos singulares tangencia as reflexões de François Recanati, um dos maiores defensores contemporâneos da metáfora dos arquivos mentais. Em artigo recente (2014, p. 165), porém, Recanati adverte que "um número crescente de filósofos tem utilizado a metáfora dos arquivos mentais para iluminar os pensamentos singulares. Pessoas diferentes, porém, elaboram a metáfora diferentemente". Esta advertência é fundamental para a apreciação da crítica de Azzouni aos arquivos mentais, pois tais críticas são endereçadas à elaboração desta metáfora por Tim Crane.

De maneira resumida, duas características diferenciam os arquivos mentais gerais e singulares, no entender de Crane. Por um lado, a maneira como o sujeito concebe a aplicação da informação armazenada no arquivo — isto é, o que tornaria geral um determinado arquivo seria o fato de o sujeito compreender uma de suas informações como sendo verdadeira quando aplicada a muitos objetos, e, inversamente, o que tornaria singular um arquivo seria o fato de o sujeito não compreender uma informação nele contida como aplicando-se a muitos objetos. Por outro lado, a maneira como uma determinada informação sobre um objeto passa a ser armazenada no arquivo — ou seja, enquanto as informações nos arquivos mentais singulares seriam passíveis de revisão, pelo menos uma informação em um arquivo mental geral seria essencial para a estruturação do mesmo. Com efeito, Azzouni questionará ambas as características.

A primeira crítica de Azzouni à elaboração de Crane consiste no fato de podermos pensar pluralidades de maneira singular. Afinal, isto embaralharia a primeira das duas características. Se podemos pensar pluralidades de maneira singular, ou, nos termos de Azzouni, se podemos desenvolver pensamentos direcionados a pluralidades de objetos, então somos capazes de compreender uma informação em um arquivo mental singular como fazendo

sentido quando aplicada a muitos objetos. Azzouni oferece como exemplos os casos de "Bourbaki" e da "família Gaona". Sobre o primeiro, ele afirma (2011, p. 49): "meu arquivo "Bourbaki" contém o fato de que houve vários deles ao longo do tempo, que nenhum deles era americano e algumas outras coisas pouco informativas (e, provavelmente, informações equivocadas também)". Já sobre o segundo (*idem*): "sou quase inteiramente ignorante dos detalhes acerca dos membros individuais desta trupe. Não tenho dúvidas, porém, de que minha sentença anterior corresponde a um pensamento direcionado para um objeto sobre eles, mesmo que eu sequer saiba quantos eles são". Esta crítica, no entanto, apenas tangencia o verdadeiro problema dos arquivos mentais.

Azzouni pensa que a metáfora dos arquivos mentais oferece uma explicação superficial para a diferença entre os pensamentos direcionados para objetos e os pensamentos descritivos. Para isso, ele supõe que a diferença entre ambas as maneiras de pensar se dá tal como pensa Crane, isto é, através da maneira como novas informações associadas com o objeto de nosso pensamento é organizada no arquivo. Assim, "quando um arquivo mental é direcionado para um objeto, qualquer uma das descrições no arquivo parece passível de revisão. […] Quando, porém, o arquivo é descritivo, certas descrições são tomadas como essenciais para o arquivo, e qualquer referência que utilize tal arquivo deve ser mediada por esta descrição" (*ibid.*, p. 52). Este quadro significa que, por exemplo, ao pensarmos singularmente sobre "Bourbaki", mesmo que venhamos a descobrir que este nome não se refere a uma pessoa específica, ainda podemos pensar sobre ele por intermédio de um arquivo mental singular.

Em contrapartida, no caso dos arquivos mentais descritivos, independentemente do que venhamos a descobrir sobre o objeto de nosso pensamento, a descrição essencial para a estruturação do arquivo nele permaneceria. O problema é que, "no pensamento, nós marcamos intuitivamente as diferenças entre os casos nos quais o insucesso da descrição definida faz com que não saibamos se algo está sendo pensado ou não; concomitantemente, se a descrição é passível de revisão ou não" (*idem*). Isto é, continua Azzouni, "ainda que

um indivíduo use um nome, nós podemos compreender uma diferença genuína em seu pensamento entre pensar com um nome *através* de uma descrição, e não pensar assim" (*idem*). E arremata:

> A possibilidade de revisão da descrição associada ao pensamento de um objeto ser o critério para saber se este pensamento é descritivo ou não faz com que o teste intuitivo para distinguirmos os pensamentos orientados para objetos de pensamentos descritivos seja experimentos de pensamento nos quais a descrição falha, mas a rota através da qual a pessoa está pensando no objeto permanece inalterada (*idem*).

Em outras palavras, se quisermos determinar a diferença entre pensamentos orientados para objetos e pensamentos descritivos, precisamos cavar um pouco mais fundo do que nos permite a metáfora dos arquivos mentais, uma vez que esta metáfora, na elaboração de Crane, quase que apenas transfere para um registro psíquico as dificuldades semânticas de se diferenciar os pensamentos em questão.

A busca pelo fundamento da distinção entre os pensamentos orientados para objetos e os pensamentos descritivos faz com que Azzouni resgate a definição da noção de *acquaintance* oferecida por Recanati. Ainda que esta noção tenha sido descartada quando se atinou à possibilidade de pensarmos singularmente objetos inexistentes, ou seja, objetos com os quais não desenvolveríamos nenhuma relação de *acquaintance*, sua definição pode auxiliar na especificação da fonte da diferença entre as duas maneiras de pensar consideradas. Afinal, a noção de *acquaintance* constitui o mais difundido ponto de partida na caracterização dos pensamentos singulares.

Para Recanati (2010, p. 152), as relações de *acquaintance* seriam relações epistemicamente recompensadoras, de modo que "pensar em um objeto direta ou não-descritivamente é nele pensar através desta relação. Neste caso, o que determina a referência — aquilo sobre o que se pensa — é a relação". Para Azzouni, apesar de esta definição indicar um bom caminho, ela demandaria um refinamento. Deste modo, ele sugere que abandonemos os adjetivos "direta" e "não-descritivamente" e descrevamos a relação epistemicamente recompensadora como "uma relação que pode ser explorada como um 'canal informacional' para o objeto em questão" (AZZOUNI, 2011, p. 53).

Obviamente, à luz dos argumentos anteriores, Azzouni também sugere que abandonemos "a ideia de que tais canais informacionais devam nos conectar com objetos atuais — que eles *devam ser* relações" (*idem*).

Pois bem, se tais canais informacionais não devem ser relações, o que eles devem ser? Azzouni responde que

> Podemos pensar neles como *maneiras* através das quais nós podemos manipular (e manipulamos) partes próximas do mundo para extrair informações — e que, às vezes, existe algo no remoto fim de uma cadeia causal que originou nossa manipulação do mundo, e, às vezes, não há nada. Portanto, assim como posso utilizar canais informacionais para descobrir coisas sobre o planeta Saturno, também posso utilizar canais informacionais para descobrir coisas sobre Mickey Mouse. Esses canais podem ser perceptuais, é claro (e, correspondentemente, podem ser alucinatórios); mas também podem envolver testemunhos ou instrumentação científica, ou a exploração inteligente de traços que os objetos fizeram no mundo (*idem*).

E acrescenta que:

> Não deveríamos pensar nos canais informacionais de forma tão restrita; na verdade, não deveríamos pensar neles como *canais* informacionais em absoluto. Ao invés disso, deveríamos pensar neles como rotas percebidas para objetos que alguém pode ou não ser capaz de explorar. Se isto está correto, então o simples fato de que o nome de alguém encontra-se em domínio público é suficiente para criar a impressão de um 'canal informacional' para a pessoa nomeada. Pois este nome não possui nenhuma propriedade externa que indique como ele se refere ao seu referente. Logo (intuitivamente), seu papel como objeto social é aquilo que possibilita sua referência — que outras pessoas o usem para se referir ao seu referente. Quando utilizamos um nome em domínio público, por conseguinte, pensamos involuntariamente de maneira direcionada para objetos; um pensamento descritivo não é possível (*ibid.*, p. 59).

De posse desta caracterização, arrisca-se uma determinação da diferença entre os pensamentos orientados para objetos e os pensamentos descritivos:

> Se um ou outro canal informacional é ele mesmo fundamental para o pensamento, então o pensamento é orientado para objeto. Se as descrições (do canal informacional) são elas mesmas centrais, então o pensamento é descritivo. A distinção, então, não é entre duas diferentes espécies de arquivos mentais, como as sugestões anteriores afirmavam, a distinção se dá na maneira como as descrições são usadas pelo indivíduo: se o indivíduo detém-se atentamente nas descrições elas mesmas, ou se ele se detém nos canais informacionais aos quais as descrições (parcialmente) se referem. A distinção entre

> orientado para objeto e descritivo, como estou sugerindo, é uma distinção entre um foco no conteúdo de uma descrição por intermédio desta descrição em oposição a um foco no conteúdo de uma descrição independentemente desta descrição (*ibid.*, p. 54).

Estamos cada vez mais perto de compreender a razão pela qual Azzouni julga ser possível pensarmos singularmente os objetos matemáticos.

Em seu raciocínio, Azzouni afirma que a mera presença de um nome em domínio público é capaz de criar a impressão de um "canal informacional" para o ser nomeado. Desta hipótese, ele deduz que, ao utilizarmos um nome, pensamos involuntariamente de maneira direcionada para objetos, pois o emprego do nome interditaria a rota para o desenvolvimento de um pensamento descritivo. Um pouco adiante, Azzouni destaca que os diagnósticos de pensamentos descritivos quase sempre podem ser modificados para pensamentos orientados para objetos, ainda que o inverso não proceda. Isto implica na tese de acordo com a qual "a fenomenologia dos pensamentos singulares mostra que eles são bastante comuns, e que (correspondentemente) pensamentos completamente gerais parecem mais raros" (*ibid.*, p. 47). E é desta fenomenologia dos pensamentos singulares que surgem duas hipóteses.

A primeira diz respeito ao instrumentalismo semântico de Kaplan. Se esta teoria defende que basta a utilização de um nome para desenvolvermos pensamentos singulares acerca do objeto nomeado, Azzouni pensa que ela está correta. Mas com uma condição: que o nome em questão já esteja em domínio público. A possibilidade de um indivíduo inventar um nome para uso privado seria insuficiente para o desenvolvimento de pensamentos orientados para objetos.

A segunda diz respeito ao caso matemático. Azzouni afirma (*ibid.*, p. 60) que "o mero uso de um nome numérico em domínio público é suficiente para ativar pensamentos direcionados para objetos mesmo sem um (genuíno) canal informacional". Isto é, o fato de os objetos matemáticos serem publicamente nomeados seria suficiente para configurar uma relação epistemicamente recompensadora através da qual podemos manipulá-los com vistas à elaboração de novas informações, independentemente das descrições utilizadas. Diz Azzouni (*idem*) que "podemos considerar isso como o tímpano do mecanismo

dos canais informacionais que desencadeia o pensamento direcionado a objetos".[7]

Após conjecturar sobre o caso matemático, Azzouni afirma (*idem*) que "podemos, se quisermos, tratar da mesma forma os bem sucedidos pensamentos orientados para objetos em relação ao inexistente. Apenas o nome, "Mickey Mouse", em geral, é suficiente para desencadear pensamentos orientados para objetos". E não é uma simples coincidência ele se valer do caso matemático como uma porta para o caso dos objetos inexistentes. Pois, se até agora acompanhamos a argumentação de Azzouni em seu artigo sobre a natureza dos pensamentos singulares — seu papel cognitivo e sua fundamentação em canais informacionais —, não seria incorreto dizer que as reflexões presentes neste artigo são anunciadas no livro *Talking About Nothing* e complementadas pelo ensaio "Mathematical Fictions". Neste ensaio, Azzouni analisa algumas ilusões cognitivas nas quais nos encontraríamos enredados. Particularmente interessante para a teoria cujos traços esboçamos é a ilusão de que a atinência de nosso pensamento deve nos conectar com um objeto existente. Para Azzouni (2015, p. 68), "nossa matemática é cheia de ficções do mesmo tipo que aparecem nas histórias que narramos"; o problema, continua, é julgarmos que as ficções servem apenas para nos entreter. A reflexão de Azzouni sobre os pensamentos singulares lança-nos em seu nominalismo, a tese de acordo com a qual "*nenhum* termo matemático refere" (*idem*). O desenvolvimento destes traços de sua teoria, porém, nos levaria longe demais.

Questões e perspectivas

É importante compreendermos que a mera afirmação de nossa capacidade de pensar singularmente objetos matemáticos pressupõe uma série de

[7] É interessante a utilização do substantivo "tímpano" (ou "enjunta" — a palavra utilizada por Azzouni é "*spandrel*"), pois este termo, originalmente arquitetônico, designa um espaço triangular geralmente encontrado em pares, entre o topo de um arco e uma moldura retangular, um espaço cujas possibilidades artísticas demoraram a ser percebidas e exploradas pelos artistas, mas que, uma vez percebidas, ampliaram e complexificaram a beleza das construções em que estavam inseridos. Azzouni parece sugerir que o mesmo acontece com os objetos matemáticos. Isto é, que os pensamentos orientados para objetos relacionados com os objetos matemáticos constituiriam um espaço teórico que sempre esteve presente, mas apenas quando demos por sua presença começamos a explorar suas mais diversas possibilidades.

respostas para uma série de questões que aqui não encontraram seu desenvolvimento. Em primeiro lugar, a existência de pensamentos singulares não é consensual no interior da tradição filosófica, fato que implicaria em uma defesa inicial da própria existência de tais pensamentos. Em segundo lugar, é possível admitir a existência de pensamentos singulares sem que, com isso, admita-se a possibilidade de assim pensarmos quaisquer objetos com os quais não estabeleçamos uma relação causal, fato que implicaria na defesa de um caráter contingente da relação causal em nosso processo de pensamento. Em terceiro lugar, é possível reconhecer que, em nosso pensamento, nós nos relacionamos singularmente com "objetos ficcionais" ou até mesmo com "objetos alucinatórios", sem que com isso queiramos nos comprometer com a existência de pensamentos singulares sobre objetos matemáticos, fato que demandaria uma caracterização mais aguda da mecânica conceitual envolvida na possibilidade de assim pensarmos os objetos matemáticos. Em verdade, as desinteligências no reino dos pensamentos singulares multiplicam-se a cada tópico destacado.

As teorias de Azzouni e Davies oferecem distintas portas de entrada não apenas para o estudo dos pensamentos singulares como para o estudo sobre pensamentos singulares sobre objetos matemáticos. Por um lado, Davies sustenta que os arquivos mentais constituem o fundamento dos pensamentos singulares, que os pensamentos singulares são caracterizados à luz de seu conteúdo e que dependem da existência de seu objeto. Em resumo, ele investiga de que maneira podemos ampliar o conjunto dos objetos singularmente pensados, de modo a nele incluirmos os objetos matemáticos, respeitando nossas intuições habituais sobre os termos matemáticos. Por outro lado, Azzouni sustenta que os pensamentos singulares deveriam, à luz de seu comportamento cognitivo, ser caracterizados como uma maneira de pensar, o que faz com que sua designação seja modificada e eles passem a ser denominados como pensamentos orientados para objetos. Este posicionamento implica que nossos pensamentos não dependem da existência do objeto, uma vez que podemos pensar em diversos objetos inexistentes. O caso dos objetos ficcionais (tais como "Mickey Mouse") é mais discutido no artigo que orientou a exposição,

mas indiquei que, no entender de Azzouni, os objetos matemáticos possuem o mesmo estatuto metafísico que os objetos ficcionais. Para onde vamos a partir daqui?

Uma maneira de concluirmos esta introdução ao estudo dos pensamentos singulares sobre objetos matemáticos consiste em decidirmo-nos por uma das alternativas apresentadas, isto é, podemos argumentar que a análise de Davies mostra-se mais promissora do que a de Azzouni, ou vice-versa. Por outro lado, podemos nos exercitar dialeticamente na elaboração de um argumento capaz de abranger as teorias de Davies e de Azzouni. Naturalmente, também podemos argumentar que ambas as teorias não acertam o alvo em questão, seja este uma caracterização mais específica dos pensamentos singulares sobre objetos matemáticos. No entanto, não gostaria de seguir por nenhum desses caminhos; gostaria apenas de destacar algumas questões suscitadas pelo estudo que encontrou neste artigo um de seus retratos.

O fato de Davies incorporar em sua proposta o quadro teórico desenhado por Jeshion direciona nosso olhar para a teoria por ela desenvolvida. Com efeito, apesar de Jeshion prometer, aqui e ali, um estudo mais detalhado para o caso dos pensamentos singulares acerca de objetos matemáticos,[8] e apesar de ela ter escrito um artigo inteiramente dedicado à noção de "intuição matemática",[9] não há nenhum desenvolvimento de seu Cognitivismo nesta direção. A análise dos pensamentos singulares sobre objetos matemáticos de uma perspectiva cognitivista seria bem-vinda no conjunto de estudos sobre a natureza de tais pensamentos.

Do mesmo modo, a advertência de Recanati lança uma sombra sobre a caracterização dos arquivos mentais efetuada por Davies.[10] Aparentemente, sua caracterização dos arquivos mentais apenas conecta algumas críticas efetuadas por Rachel Goodman com a configuração conceitual de Recanati. O problema, aqui, consiste em saber se haveria algum denominador comum capaz de reunir os defensores da teoria dos arquivos mentais (tais como Recanati, Jeshion e Crane), ou se, como afirma Recanati, cada pessoa efetivamente

[8] Cf. Jeshion 2002, p. 56-57; Jeshion 2004, p. 607-610.
[9] Cf. Jeshion 2014b.
[10] Cf. Recanati, 2014, p. 165.

elabora a metáfora dos arquivos mentais à sua maneira. Na recém-publicada coletânea *Singular Thought and Mental Files*, Murez, Smortchkova e Strickland analisam o que eles consideram a mais ambiciosa versão da teoria dos arquivos mentais vinculada aos pensamentos singulares.[11] Seu trabalho, porém, tem por objetivo avaliar criticamente o modo pelo qual diversos filósofos teriam superestimado a noção de "arquivo" utilizada nas ciências cognitivas, desenvolvendo uma noção de arquivo mental que não encontraria sustentação nas pesquisas psicológicas (algo já indicado por Laura Schroeter em relação à "transparência" de um arquivo mental para o sujeito cognoscente). A investigação das propriedades consensuais dos arquivos mentais também seria bem-vinda nos estudos dos pensamentos singulares.

Por sua vez, Azzouni, em sua teoria, claramente favorece o papel desempenhado pelos nomes próprios em relação aos demais termos singulares. Em verdade, à exceção dos nomes descritivos e indexicais, podemos dizer que praticamente todas pesquisas sobre os pensamentos singulares privilegiam ora os nomes próprios, ora os pronomes demonstrativos. Deste modo, podemos perguntar: haveria uma hierarquia entre os termos singulares? Isto é, seriam os nomes próprios mais fundamentais do que os pronomes demonstrativos ou vice-versa? Assim como Azzouni, Jeshion também concede uma primazia aos nomes próprios, ao passo que Recanati privilegia o papel dos pronomes demonstrativos.[12] Recentemente, Imogen Dickie dedicou um longo ensaio à compreensão dos termos singulares em nossa linguagem natural a partir de sua teoria do "foco cognitivo".[13]

Ainda no início de seu artigo, Azzouni afirma ser uma questão delicada especificar qual seria a relação entre pensamento e linguagem — um diagnóstico que, por extensão, se aplicaria à relação entre pensamento singular e proposição singular. De fato, esta é uma questão bastante sútil. Por um lado, Elizabeth Spelke e Susan Hespos assinalam, a partir de seus experimentos, que as crianças nascem com um sistema independente da linguagem que as auxilia a pensar nos objetos; por outro lado, Lera Boroditsky e Guy Deutscher

[11] Cf. Murez, Smortchkova, and Strickland 2020.

[12] Cf. Jeshion 2009; Jeshion 2010c; Jeshion 2014; Recanati 2010; Recanati 2012

[13] Cf. Dickie 2020.

assinalam que a linguagem é a responsável por moldar nossos pensamentos, cuja implicação mais imediata consiste na hipótese segundo a qual elaboramos pensamentos diferentes em diferentes línguas.[14] A grandeza, a complexidade e a importância deste problema para os estudos sobre os pensamentos singulares é bastante perceptível.

Obviamente, muitos outros problemas poderiam ser destacados. Estas questões, porém, constituem o núcleo das pesquisas que procuram averiguar a fertilidade da seguinte hipótese: ao especificarmos os mecanismos favoráveis ao desenvolvimento de pensamentos singulares, particularmente no que diz respeito ao caso dos objetos matemáticos, esquadrinhamos o funcionamento psico-semântico dos pensamentos singulares em seu contraste com os pensamentos descritivos. Contudo, estaríamos com isso estabelecendo uma espécie de pêndulo cognitivo do ser humano? Isto é, será que estaríamos destinados a pensar apenas de maneira descritiva ou de maneira singular? Não haveria aqui possibilidades ainda inexploradas? Conseguiríamos nos relacionar com os mais variados objetos de uma maneira que esteja para além das formas condensadas pelos pensamentos singulares e descritivos? Apesar de não possuir uma resposta para estas perguntas, elas ajudam-me a colocar em perspectiva todos os argumentos aqui estudados.

Referências

AZZOUNI, Jody. "Empty De Re Attitudes About Numbers". *Philosophia Mathematica*, v. 3, n. 17, 2009, p. 163-188.

———————. *Talking About Nothing: Numbers, Hallucinations, and Fictions*. Oxford: Oxford University Press, 2010.

———————. "Singular Thoughts (Objects-Directed Thoughts)". *Proceedings of the Aristotelian Society Supplementary Volume*, v. 85, 2011, p. 45-61.

———————. "Mathematical Fictions". In: SUKLA, Ananta (ed.). *Fiction and Art: explorations in contemporary philosophy*. London: Bloomsbury, 2015, p. 63-77.

BACH, Kent. *Thought and Reference*. Oxford: Clarendon Press, 1987.

[14] Cf. Hespos, and Spelke 2004; Boroditsky 2001; Deutscher 2010.

———————. “Getting a Thing Into a Thought”. In: JESHION, Robin (ed.). *New Essays on Singular Thought*. Oxford: Oxford University Press, 2010, p. 39-63.

BORODITSKY, Lera. “Does Language Shape Thought? Mandarin and English Speakers’ Conceptions of Time”. *Cognitive Psychology*, v. 43, 2001, p. 1-22.

CUNHA, Celso, e CINTRA, Lindley. *Nova Gramática do Português Contemporâneo*. Rio de Janeiro: Editora Lexikon, 2008.

CRANE, Tim. “The Singularity of Singular Thought”. *Proceedings of the Aristotelian Society Supplementary Volume*, v. 85, 2011, p. 21-43.

———————. *The Objects of Thought*. Oxford: Oxford University Press, 2013.

DAVIES, James E. “Towards a theory of singular thought about abstract mathematical objects”. *Synthese*, v. 196, n. 10, 2019, p. 1-24.

DEUTSCHER, Guy. *Through the language glass: why does the world looks different in other languages*. New York: Metropolitan Books, 2010.

DICKIE, Imogen. “Understanding Singular Terms”. *Proceedings of the Aristotelian Society Supplementary Volume,* v. 94, 2020, p. 19-55.

GOODMAN, Rachel. “Against the mental files conception of singular thought”. *Review of Philosophy and Psychology*, v. 7, 2016, p. 1–25.

HESPOS, Susan, e SPELKE, Elizabeth. “Conceptual precursors to language”. *Nature*, v. 430, 2004, p. 453-456.

JESHION, Robin. “Acquaintanceless *De Re* Belief”. In: CAMPBELL, O’ROURKE, e SHIER (eds.). *Meaning and Truth*. New York: Seven Bridges Press, 2002, p. 53-74.

———————. “Descriptive Descriptive Names”. In: BEZUIDENHOUT, Anne, e REIMER, Marga (eds.). *Descriptions and Beyond*, Oxford: Clarendon Press, 2004, p. 591-612.

———————. “The Significance of Names”. *Mind and Language*, v. 24, n. 4, 2009, p. 370-403.

——————— (ed.). *New Essays on Singular Thought*. Oxford: Oxford University Press, 2010a.

———————. “Introduction to New Essays on Singular Thought”. In: JESHION, Robin (ed.). *New Essays on Singular Thought*. Oxford: Oxford University Press, 2010b, p. 1-35.

———————. “Singular Thought: Acquaintance, Semantic Instrumentalism, and Cognitivism”. In: JESHION, Robin (ed.). *New Essays on Singular Thought*. Oxford: Oxford University Press, 2010c, p. 105-140.

———————. “Two Dogmas of Russellianism”. In: GARCÍA-CARPINTERO, Manuel, and MARTÍ, Genoveva (eds.). *Empty Representations: Reference and Non-Existence*. Oxford: Oxford University Press, 2014a, p. 67-90.

———————. “Intuiting the Infinite”. *Philosophical Studies*, v. 171, 2014b, p. 327-349.

McDOWELL, John. *Meaning, Knowledge, and Reality*. Cambridge: Harvard University Press, 1998.

MUREZ, SMORTCHKOVA, e STRICKLAND. “The Mental Files Theory of Singular Thought”. In: GOODMAN, GENONE, e KROLL (eds.). *Singular Thought and Mental Files*. Oxford: Oxford University Press, 2020, 107-142.

RECANATI, François. “Singular Thought: In Defense of Acquaintance”. In: JESHION, Robin (ed.). *New Essays on Singular Thought*. Oxford: Oxford University Press, 2010, p. 141-189.

———————. *Mental Files*. Oxford: Oxford University Press, 2012.

———————. “Empty Singular Terms in the Mental File Framework” . In: GARCÍA-CARPINTERO, Manuel, and MARTÍ, Genoveva (eds.). *Empty Representations: Reference and Non-Existence*. Oxford: Oxford University Press, 2014, p. 165-182.

———————. *Mental Files in Flux*. Oxford: Oxford University Press, 2016.

SCHROETER, Laura. “The Illusion of Transparency”. *Australasian Journal of Philosophy*. v. 85, n. 4, 2007, p. 597-618.

O NONSENSE DE LEWIS CARROLL

John Lennon Lindemann[1]

RESUMO

O presente trabalho apresenta uma noção de *nonsense* própria às obras de Lewis Carroll e elaborada com base em seu próprio arcabouço teórico, seguida de uma análise das relações entre as noções de *nonsense* de Lewis Carroll e Wittgenstein, concluindo que ambos tiveram atitudes análogas em relação ao *nonsense*, usando-o como um meio para induzir ideias instrutivas nos seus leitores.

Palavras-chave: Nonsense. Lewis Carroll. Wittgenstein. Lógica silogística.

ABSTRACT

The present work presents a notion of nonsense specific to Lewis Carroll's work and elaborated with base on his own theoretical framework, followed by an analysis of the relations between Carroll's and Wittgenstein's notions of nonsense, concluding that both had similar attitudes towards nonsense, using it as a tool to induce instructive ideas to their readers.

Keywords: Nonsense. Lewis Carroll. Wittgenstein. Syllogistic logic.

Introdução

Lewis Carroll, pseudônimo de Charles Lutwidge Dodgson (1832-1898), é afamado por seus livros *nonsense*, "clássicos da literatura inglesa" (GARDNER, 2002b, p. xvii). Além de escritor, Charles Dodgson foi professor de matemática no Christ College, em Oxford, e ele também é reconhecido, mediante seu pseudônimo, por suas contribuições ao desenvolvimento da lógica, especialmente por seu método diagramático de resolução de silogismos apto ao tratamento de termos negativos (ABELES, 2007). Cabe ressaltar que o *nonsense* carrolliano não tem o seu *locus* restrito às suas obras literárias, mas também ocorre em exemplos nos seus trabalhos sobre lógica.

[1] Mestre em Filosofia, doutorando em Filosofia pela UFSM. Bolsista Capes.
E-mail: johnllindemann@hotmail.com. ORCID: 0000-0001-8613-2132.

A noção de *nonsense* já recebeu tratamento de muitos filósofos ilustres (HUSSERL, 1900; DELEUZE, 1974), entre os quais Ludwig Wittgenstein (1889-1951) se destaca pela influência que sua abordagem exerceu na história da filosofia. Wittgenstein define enunciados *nonsense* como destituídos de valor veritativo: "[…] uma combinação de palavras é nonsense quando não pode ser entendida, porque nenhum sentido é ou pode (exceto trivialmente) ser reconhecido" (PITCHER, 1965, p. 591, tradução nossa).

O presente trabalho defende uma noção de *nonsense* própria às obras de Carroll e oriunda do arcabouço teórico de seus trabalhos acadêmicos, seguida da análise das relações que tal noção mantém com a noção de *nonsense* wittgensteiniana.

Já há trabalhos anteriores dedicados ao estabelecimento de uma noção de *nonsense* própria às obras de Carroll (MONTOITO, 2019; DELEUZE, 1974), mas ainda não há uma apresentação com base nas posições teóricas sobre lógica do próprio autor, o que é relevante devido às suas contribuições para o desenvolvimento da lógica.

Por sua vez, o trabalho mais importante sobre a relação entre o *nonsense* de Carroll e Wittgenstein se chama *Wittgenstein, Nonsense, and Lewis Carroll*, de Pitcher (1965) – defendendo que as atitudes de Carroll e Wittgenstein em relação ao *nonsense* são radicalmente distintas (PITCHER, 1965, p. 611), tese que será atacada no presente trabalho.

Guiado por tais objetivos, o presente trabalho se divide em três seções. A primeira seção apresenta uma noção de *nonsense* própria às obras de Carroll e elaborada com base nas posições teóricas em lógica do autor, a segunda seção apresenta a noção de *nonsense* wittgensteiniana que será utilizada na terceira seção, onde se apresenta uma análise das relações entre as noções de *nonsense* de Carroll e Wittgenstein e se conclui que os dois autores têm atitudes análogas em relação ao *nonsense*, usando-o como um meio de induzir instrução nos seus leitores.

1. O *Nonsense* Carrolliano

O *nonsense* de Lewis Carroll é comumente categorizado como a expressão mais relevante de um gênero literário inaugurado por Edward Lear (1812-1888), seu contemporâneo, mas que pode ser associado com muitos autores até os dias de hoje. Em casos mais restritos, como em Reichert (1974), Carroll e Lear são considerados os únicos autores associados ao gênero, que terminou com a morte de Carroll: "Reichert identifica um traço comum ao nonsense e ao Zeitgeist vitoriano, que seria a base da estreita relação entre esse fenômeno e o período em que ele se deu" (ÁVILA, 1996, p. 20).

Independente da relação que o *nonsense* carrolliano possa manter com as obras de Lear e dos demais autores associados a esse gênero literário, a definição de uma noção de *nonsense* específica às obras de Carroll também foi objeto de muitos trabalhos, com destaque para Montoito (2019), que, em congruência com considerações de Deleuze (1974), defende uma noção segundo a qual o *nonsense* carrolliano não é avesso ao sentido, mas nasce justamente do exercício da produção de novos sentidos – possível graças aos seus criativos universos de ficção (MONTOITO, 2019, p. 37).

Assumindo a tese de Deleuze-Montoito como pano de fundo, a presente seção dedica-se à análise do *nonsense* de Carroll, sendo pautada pela perspectiva do próprio autor, apresentando o contexto histórico de sua constituição e examinando os pormenores de suas características por meio do arcabouço teórico de seus trabalhos acadêmicos.

1.1. A Constituição Histórica do Nonsense Carrolliano

O *nonsense* literário nasce na Era Vitoriana, com a publicação de *A Book of Nonsense*, de Lear (1846), mas é graças a Carroll que ele se torna um fenômeno mundial. Inclusive o poema "Pargarávio"[2], que aparece em *Através do Espelho e o que Alice encontrou por lá* (CARROLL, 2002, p. 143-144), é considerado como o exemplo de literatura *nonsense* por excelência (MALCOLM, 1997, p. 14).

[2] Do original "Jabberwocky" (CARROLL, 1872, p. 21-22), mais conhecido no Brasil como "Jaguadarte" devido à célebre tradução de Augusto de Campos (CARROLL, 2014).

Tal gênese não pode ser compreendida fora do contexto histórico do qual emergiu. Segundo Gardner (2002a, p. vii):

> No caso de Alice, estamos lidando com uma espécie de nonsense muito curioso, complicado, escrito para leitores britânicos de um outro século, e precisamos conhecer um grande número de coisas que não fazem parte do texto se quisermos apreender todo o seu espírito e sabor. É até mais grave que isso, porque algumas piadas de Carroll só podiam ser compreendidas por quem residia em Oxford, […].

Apesar do rigor moral e das hierarquias sociais fortemente estabelecidas, características representadas inclusive na própria Rainha Vitória (1819-1901), a Era Vitoriana também foi um período de muitas transformações e invenções, por exemplo, a invenção da fotografia, arte da qual Carroll foi um entusiasta, inclusive "chegou a ser reconhecido como o melhor fotógrafo de crianças do século XIX" (COHEN, 1998, p. 16).

Sobre os hábitos culturais do período, segundo Ávila (1996, p. 20):

> A cultura do lazer [...] é marcada pela crescente popularidade dos jogos com palavras, charadas e palavras-cruzadas, em que as palavras se relacionam umas com as outras por critérios que não levam em conta seu significado. É notável, também, a multiplicação dos museus (nos quais os objetos mais diversos são colocados lado a lado sob o rótulo comum de "peça de exposição"), e dos dicionários nos quais a ordem alfabética tem prioridade sobre as relações semânticas […].

Carroll também foi um entusiasta na criação de jogos e os "jogos de palavras sempre foram seus favoritos [...]" (SHAWYER, 1953, p. xxv), com destaque para a criação do jogo de palavras chamado "Doublets"[3] (DODGSON, 1880), cujo objetivo é transformar uma palavra em outra seguindo regras que permitem alterar as letras. Ele também criou inúmeros enigmas e desafios lógicos, muitos dos quais foram compilados por Abeles em *The Pamphlets of Lewis Carroll* (CARROLL, 2010).

Em relação à história do desenvolvimento da lógica, a Era Vitoriana também é o marco de um período especial. Superando o legado de Aristóte-

[3] Miller (2000) e Mortari (2001, p. 233-234) apresentam o jogo Doublets como ferramenta didática para o ensino de Dedução Natural: "Uma maneira de diminuir a dor da iniciação [de manipulações formais] é começar com jogos de palavras, dos quais o Doublets de Lewis Carroll é um exemplo familiar" (MILLER, 2000, p. 1, tradução nossa).

les, George Boole inaugura uma nova era da lógica ao apresentar formas de raciocínio válido não limitadas às formas da silogística aristotélica. Segundo Quine (1996, p. 15):

> [...], a lógica sofreu tal evolução que pode ser considerada uma ciência nova. Essa evolução é considerada como tendo seu início nas pesquisas ainda rudimentares do matemático George Boole, em meados do século passado [XIX].

Tal período, no qual as obras lógicas de Carroll e de seus pares se situa, é conhecido como o período da álgebra da lógica, que começa com as inovações de Boole e se estende até a popularização do cálculo de predicados criado por Frege (1879), que "constitui desde então a base da lógica moderna" (KENNY, 1999, p. 438).

Diante de tantas invenções e da miscelânea de novidades que não se limita aos objetos materiais, mas se estende às novas formas de raciocinar e de se relacionar com a linguagem, nasce o *nonsense* carrolliano: uma resposta às novas possibilidades do mundo, um fruto direto das novas formas de raciocinar e das novas maneiras de usar e subverter a linguagem.

Deve-se deixar claro, como adverte Montoito (2019, p. 37), que o *nonsense* carrolliano não se confunde com o uso ordinário do termo, que se refere ao absurdo e à inconsistência lógica, mas faz uma apologia à lógica e desbrava os seus distintos usos. Embora o País das Maravilhas não seja submisso à nossa realidade, ele se curva diante das relações lógicas que seus elementos mantêm entre si, de tal modo que o *nonsense* não é avesso ao sentido, mas nasce justamente da formação de um novo sentido sob outro sistema referencial.

Um exemplo pode ser encontrado na constante interação de Alice com objetos que modificam o seu tamanho (CARROLL, 2002, p. 18):

> Pouco depois deu com os olhos numa caixinha de vidro debaixo da mesa: abriu-a, e encontrou um bolo muito pequeno, com as palavras "COMA-ME" lindamente escritas com passas sobre ele. "Bem, vou comê-lo", disse Alice; "se me fizer crescer, posso alcançar a chave; se me fizer diminuir, posso me esgueirar por baixo da porta; assim, de uma maneira ou de outra vou conseguir chegar ao jardim; para mim tanto faz!"

Na sequência dos eventos, o bolo faz com que Alice espiche ao terminar de comê-lo, efeito antecipado por ela, mostrando como já estava se habituando aos objetos do universo *nonsense*, interagindo com eles supondo seus próprios atributos. O raciocínio de Alice pode ser formalizado utilizando a linguagem do Cálculo Proposicional Clássico tal como se segue, onde "P" representa "Alice come o bolo", "Q" representa "Alice aumenta de tamanho", "R" representa "Alice diminuí de tamanho", "S" representa "Alice alcança a chave da porta", "T" representa "Alice se esgueira por baixo da porta" e "U" representa "Alice chega ao jardim":

$$(P \rightarrow (Q \vee R)), (Q \rightarrow S), (R \rightarrow T), ((T \vee S) \rightarrow U) \therefore (P \rightarrow U)$$[4]

O argumento acima é tautologicamente válido, demonstrando o rigor lógico presente no raciocínio de Alice mesmo quando aplicado às suas interações absurdas com objetos do universo *nonsense*.

É justamente a partir da possibilidade de novos sentidos, respectivos a outros sistemas referenciais, que somos apresentados às situações mais engraçadas da literatura *nonsense* carrolliana: seja fruto da ambiguidade de um termo, como a intenção do Camundongo de secar a todos com uma conversa seca (CARROLL, 2002, p. 27-28)[5], da transformação do sentido dos termos, como o faz Humpty Dumpty ao atribuir às palavras o significado arbitrário que deseja (CARROLL, 2002, p. 199-211)[6] ou dos muitos neologismos criados por Carroll.

Entre uma vasta gama de neologismos, cito "sillygism", uma variante do termo "silogismo" criada por Carroll (2005, p. 148) e traduzida por Montoito (2019, p. 27) como "tontogismo", com a ressalva de que o termo "tonto" não é usado como pejorativo, mas com humor, em uma roupagem carinhosa.

[4] Essa análise lógica foi comunicada a mim pelo Prof. Dr. Frank Thomas Sautter.

[5] Alice (que havia encolhido) acompanhava um grupo de pequenas aves e animais, todos estavam molhados pois haviam acabado de sair do lago salgado criado pelas lágrimas derramadas por Alice (quando estava grande). Infelizmente, o plano do Camundongo fracassou e sua conversa seca não os secou (CARROLL, 2002, p. 23-28).

[6] Segundo Gardner (2002c, p. 205), Carroll estava ciente da riqueza da tese semântica de Humpty Dumpty: "Humpty adota o ponto de vista conhecido na Idade Média como nominalismo; a ideia de que termos universais não se referem a existências objetivas, nada mais sendo que flatus vocis, emissões verbais".

Montoito se apropria do termo "tontogismo" como uma categorização para pensar a lógica presente nas obras de Carroll[7], em suas palavras:

> [...], percebo no tontogismo uma manipulação intencional e criativa da lógica formal que, apoiada nos usos – desusos e abusos – da linguagem, comunica ideias lógicas. (MONTOITO, 2019, p. 27)
> O tontogismo é um buraco de fechadura através do qual, estando no mundo real, espia-se a lógica de um universo fantasioso em que tudo pode mudar ou deixar de funcionar, menos a própria lógica; ou, estando no mundo imaginário, se espia o mundo real para entendê-lo através de outro sistema de referências. (MONTOITO, 2019, p. 28)

O termo "tontogismo" é utilizado por Montoito para caracterizar a lógica nas obras de Carroll em um sentido amplo, não restrito às suas publicações literárias. Embora Charles Lutwidge Dodgson tenha criado o seu pseudônimo "baseado na inversão latinizada de seus dois primeiros nomes" (COHEN, 1998, p. 99) especificamente para suas obras literárias, de tal modo que não fossem relacionadas às suas publicações de caráter acadêmico, Montoito defende que não há tal ruptura, mas apenas um descompasso proposital, "ora mantendo-se mais formal, ora dando mais vazão ao nonsense, sem jamais impor que um estilo anule o outro" (2019, p. 54).

Observando as publicações acadêmicas e literárias de Carroll como partes de um mesmo corpo, percebe-se a clara ocorrência de seu estilo *nonsense* em seus trabalhos sobre lógica e filosofia da lógica, sendo amplamente utilizado nos exemplos de seus manuais de lógica – estratégia com grande apelo didático, mas também ocorrendo no importante artigo *What the Tortoise Said to Achilles* (CARROLL, 1895), que problematiza os fundamentos da lógica.

Por exemplo, as premissas de "O Problema dos Porcos e Balões"[8], um exemplo de sorites[9] presente em *Symbolic Logic*, obra fruto de seus estudos

[7] Montoito (2019) usa a expressão "lógica do nonsense" para designar a lógica presente nas obras de Carroll. A expressão foi utilizada originalmente em sua dissertação de mestrado (2007). O termo "lógica" ocorre de modo não técnico na expressão, não designando uma nova classe de lógica, mas o modo pelo qual Carroll apresenta a lógica clássica em seus trabalhos.

[8] "O Problema dos Porcos e Balões" é examinado em Lindemann (2020).

[9] *Lato sensu*, sorites é um encadeamento de silogismos.

mais maduros e na qual encontramos as suas maiores contribuições no desenvolvimento da lógica (CARROLL, 1986, p. 378-380, tradução nossa):

> 1. Todos, que nem dançam na corda bamba nem comem brioches, são velhos.
> 2. Os porcos, suscetíveis à vertigem, são tratados com respeito.
> 3. Um balonista sensato leva um guarda-chuva com ele.
> 4. Ninguém, que deve almoçar em público, parece ridículo e come brioches.
> 5. As criaturas jovens, que sobem em balões, estão sujeitas à vertigem.
> 6. Criaturas gordas, que parecem ridículas, podem almoçar em público, desde que elas não dancem na corda bamba.
> 7. Nenhuma criatura sábia dança na corda bamba, se for suscetível à vertigem.
> 8. Um porco parece ridículo, carregando um guarda-chuva.
> 9. Todos, que não dançam na corda bamba, e que são tratados com respeito, são gordos.

Dada as nove premissas acima, Carroll (1986, p. 378-380) demonstra a validade da conclusão "Nenhum porco sábio e jovem é balonista" por meio do seu método de sublinhar (CARROLL, 1986, p. 138-139), um método de raciocínio linear em que cada proposição é traduzida para uma fórmula na qual os termos em língua natural e a relação lógica que mantém entre si são representadas de maneira simbólica e, partindo de qualquer par de fórmulas, uma terceira fórmula é deduzida segundo certas regras de inferência, repetindo o procedimento até que se obtenha a fórmula que representa a conclusão almejada.

O método de sublinhar exemplifica como o trabalho de Carroll estava em sincronia com os principais desenvolvimentos lógicos de sua época. Segundo Valencia (2004, p. 389), a principal característica do período da álgebra da lógica consiste em traduzir um dado conteúdo para análise lógica em uma forma simbólica, aplicar técnicas para a sua resolução e traduzir o resultado de volta para a linguagem original, tal como no método carrolliano.

Em todos os exemplos *nonsense* nos trabalhos lógicos de Carroll, independente do quão incomum seja o conteúdo expresso por suas premissas, tal como no exemplo supracitado, assumir que sejam verdadeiras em um sistema de referências alternativo implica assumir a verdade da conclusão em tal sistema, desde que seja uma conclusão válida – obtida com rigor lógico.

Segue-se que não há distinção entre o *nonsense* literário e o *nonsense* encontrado em suas obras lógicas, ambos emergem da formação de um sistema de referências alternativo, no qual todas as coisas podem ser diferentes, exceto o rigor às relações lógicas que elas mantêm entre si. Assim, cada exemplo *nonsense* nas obras lógicas de Carroll apresenta um micro universo *nonsense*, completamente análogo aos extraordinários universos nonsense de suas obras literárias.

1.2. A Constituição Lógica do Nonsense Carrolliano

Se não há nenhuma diferença de espécie entre o *nonsense* literário e o *nonsense* em suas obras lógicas, então as posições teóricas de Carroll sobre lógica podem fornecer elementos que contribuam para o exame de seu *nonsense* em sentido amplo.

1.2.1 A lógica associada ao nonsense

A contribuição teórica mais popular de Carroll foi o seu método diagramático apto à resolução de silogismos com termos negativos[10]. O método foi apresentado originalmente em *The Game of Logic*, publicado em 1886, recebendo uma exposição mais detalhada em *Symbolic Logic* (CARROLL, 1986).

A teoria silogística carrolliana subjacente ao seu método diagramático é uma extensão conservativa da silogística aristotélica (LINDEMANN, 2017, p. 109-111), ou seja, a silogística aristotélica é uma subteoria da silogística carrolliana, mas a silogística carrolliana não prova nenhum novo teorema se limitada à linguagem da silogística aristotélica, isto é, sem o uso de termos negativos.

Carroll caracteriza os termos de um ponto de vista extensional, introduzindo termos negativos com o uso da noção de universo do discurso, limi-

[10] John Neville Keynes também apresenta um tratamento diagramático para termos negativos em *Studies and Exercises in Formal Logic* (1906). Embora a obra de Keynes tenha sido publicada originalmente em 1884, dois anos antes de Carroll apresentar o seu tratamento para termos negativos (1886), o método diagramático keynesiano para termos negativos só foi introduzido na quarta edição de sua obra, em 1906, vinte anos depois da inovação carrolliana.

tando a interpretação de um termo negativo ao complemento da extensão de seu respectivo termo positivo até os limites impostos pelo universo do discurso. Os termos positivo e negativo que juntos esgotam a extensão do universo do discurso são chamados de termos codivisionais (CARROLL, 1986, p. 62).

Por exemplo, supondo "Animais" como universo do discurso, o termo negativo "não-humano" denota apenas coisas que são animais e não são humanas, como coelhos e vacas. Sem o limite semântico imposto pelo universo do discurso, a interpretação do termo "não-humano" seria indefinida, pois pedras e softwares também seriam instâncias de "não-humano", problema que já havia sido diagnosticado por Aristóteles (1985, p. 141).

A noção de universo do discurso era nova no período de Carroll, recentemente introduzida por George Boole na obra *An Investigation of The Laws of Thought* (1854, p. 42, tradução nossa):

> Em todo discurso, seja da mente conversando com seus próprios pensamentos ou do indivíduo se relacionando com os outros, existe um limite assumido ou expresso dentro do qual os sujeitos de tal operação são confinados. [...], qualquer que seja a extensão do campo em que todos os objetos de nosso discurso são encontrados, esse campo pode ser adequadamente denominado de universo do discurso.

Em sua silogística, Carroll (1986, p. 107) define Universo do Discurso[11] como o Gênero ao qual os seis termos de um silogismo são Espécies, isto é, os três pares de termos codivisionais são, necessariamente, Espécies de um mesmo Gênero –que determina os limites do Universo do Discurso de cada silogismo.

Em linhas gerais, um Gênero é compreendido como uma Classe de Coisas com um certo Adjunto[12] específico, enquanto cada Espécie é uma subclasse de tal Gênero, cujos membros, além de possuírem o Adjunto que os caracteriza como membros do Gênero, também possuem (positivos) ou não possuem (negativos) algum outro Adjunto – que os distingue enquanto Espé-

[11] Carroll (1986) indica termos técnicos com inicial maiúscula. A mesma tipografia será mantida para os seus termos neste trabalho.

[12] Relativo às ocorrências de "Atributo" e "Conjunto de Atributos" em *The Game of Logic* (CARROLL, 1886), o termo "Adjunto" é introduzido apenas em *Symbolic Logic*: "Qualquer Atributo, ou qualquer Conjunto de Atributos, pode ser chamado de Adjunto" (CARROLL, 1986, p. 58, tradução nossa).

cie. *Stricto sensu*, todo termo é composto por um Nome que representa uma subclasse de membros do Universo do Discurso que possuem ou não algum Adjunto específico que os distingue de todos os outros membros do Universo do Discurso subjacente ao silogismo no qual ocorre (CARROLL, 1986, p. 59-65).

Entender as relações que tal abordagem mantém com a noção de *non-sense* carrolliana exige uma apresentação paulatina de outras posições teóricas do autor, a fim de que seus pormenores sejam examinados.

Carroll abre a exposição de suas duas principais obras lógicas por meio das mesmas considerações ontológicas, determinando o domínio de aplicação da sua teoria silogística: "O Universo contém Coisas" e "Coisas têm Atributos" (1986, p. 59, tradução nossa)[13].

Trata-se de uma ontologia hierarquicamente organizada, estabelecendo a categoria "Atributo" sob um vínculo de dependência da categoria "Coisas", sua categoria ontológica primária e fundamental. Nas palavras de Carroll (1886, p. 2, tradução nossa):

> As pessoas têm feito a pergunta "Uma Coisa pode existir sem que quaisquer Atributos lhe pertençam?" Essa questão é um grande enigma, e não vou tentar respondê-la: arrebitemos nossos narizes, e a tratemos com um insolente silêncio, como se ela não fosse digna de consideração. Mas, se colocando de outro modo, perguntarem "Um Atributo pode existir sem qualquer Coisa à qual pertença?", podemos responder imediatamente: "Não: Não mais do que um bebê pode viajar de trem sem que ninguém cuide dele!" Você nunca viu "bonito" flutuando no ar, ou espalhado no chão, sem que alguma Coisa fosse bonita, ou você viu?

Acerca da primeira questão, a qual Carroll suspende o juízo, muitos filósofos respondem negativamente, assumindo que alguns atributos são essenciais à existência de seus portadores. Loux (1998, p. 120) apresenta um exemplo que explica tal posição:

> Sócrates é também um sujeito do tipo ser humano. É Sócrates, e não algum constituinte seu, que é a coisa que é humana. Mas o tipo ser humano é o que caracteriza Sócrates como o que ele é, de tal

[13] "O mundo contém muitas Coisas" em *The Game of Logic* (CARROLL, 1886, p. 2, tradução nossa).

> maneira que neste caso nosso sujeito não é algo com uma identidade independente do universal do qual ele é sujeito. Extraindo-se de Sócrates o ser homem não sobra nada que possa ser sujeito de algo.

Sobre a segundo questão, é fato que Carroll foi um grande admirador de Aristóteles, inclusive dedicando *Symbolic Logic* (1986) à sua memória, e sua posição ao defender a impossibilidade de que um Atributo exista sem qualquer Coisa a qual pertença sugere que tal admiração talvez não seja limitada à lógica: embora não explicite sua adesão ao aristotelismo em metafísica, demonstra clara aversão ao platonismo[14].

A fim de comparação, as considerações ontológicas subjacentes à silogística aristotélica podem ser apresentadas com o quadrado ontológico (ANGELELLI, 1967, p. 12-13), obtido por meio de duas dicotomias: dividindo os entes quanto ao ser dito (universal) e não ser dito (particular) de um sujeito, assim como quanto ao estar (acidente) e não estar (substância) em um sujeito.

Tabela 1 – Quadrado ontológico

	Está em um sujeito	*Não está em um sujeito*
Dito de um sujeito	Substância universal Ex.: homem **(REGIÃO 1)**	Acidente universal Ex.: barba **(REGIÃO 2)**
Não dito de um sujeito	Substância particular Ex.: Platão **(REGIÃO 3)**	Acidente particular Ex.: barba de Platão **(REGIÃO 4)**

Fonte: Elaborado pelo autor.

Se comparado à silogística aristotélica, sob os parâmetros do quadrado ontológico, as considerações ontológicas de Carroll consideram apenas a dicotomia entre o estar e não estar em um sujeito, onde "Coisa" corresponde a "substância" e "Adjunto" corresponde a "acidente", mas, diferente de Aristó-

[14] Segundo Balaguer (2016, p. 1), compreende-se por "platonismo" a visão segundo a qual existem objetos abstratos, isto é, existem objetos inteiramente não físicos e não mentais. Ou seja, sob um léxico carrolliano e conforme a concepção platônica, o Atributo "Beleza" pode existir mesmo sem que alguma Coisa seja bonita.

teles, Carroll admite o uso de termos singulares, de tal modo que Coisas podem ocupar a Região 1 e 3 do quadrado.

Carroll também caracteriza os Adjuntos como o "que é 'dito pertencer a' uma coisa" em *The Game of Logic* (1886, p. 2, tradução nossa), reconhecendo a outra dicotomia e especificando que Adjuntos correspondem apenas à Região 2 do quadrado ontológico.

Corroborando com tal interpretação, Lindemann (2017, p. 47-48) analisa todos os exemplos apresentados por Carroll em *The Game of Logic* (1886) e *Symbolic Logic* (1986), constatando que são associáveis apenas às regiões 1, 2 e 3 do quadrado ontológico de Aristóteles.

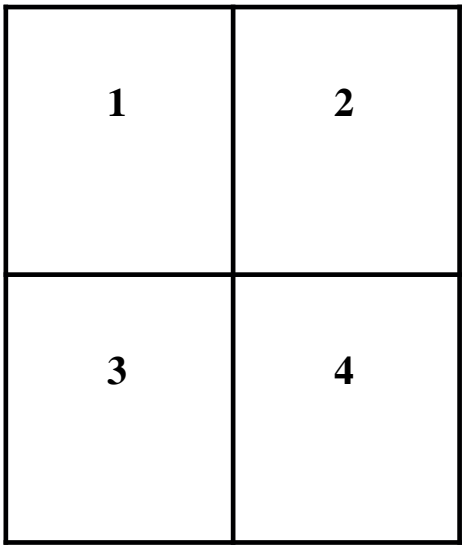
Após apresentar suas considerações ontológicas, Carroll introduz a Classificação, o processo pelo qual Coisas são agrupadas a partir de seus Atributos para a formação de Classes (CARROLL, 1986, p. 60-61). Tal processo é muito similar à "Árvore de Porfírio", estrutura lógica que apresenta a teoria aristotélica da subordinação de conceitos (PORFÍRIO, 2002).

O processo de Classificação começa supondo a Classe "Coisas"[15], que contém todas as coisas do Universo. Em um processo dicotômico, duas novas Classes são formadas a partir da determinação de um Adjunto, dividindo a Classe "Coisas" em duas subclasses, a saber, a Classe cujos membros possuem tal adjunto e a Classe cujos membros não o possuem. Qualquer Classe formada a partir da Classe "Coisas" serve como Gênero para a formação de Classes-Espécie a partir do mesmo processo dicotômico.

A relação entre os membros de quatro ou mais Classes que são Espécies de um mesmo Gênero é apresentada de maneira mais clara por meio de sua representação diagramática:

[15] A Classe "Coisas" é análoga ao conceito de "Substância" da Árvore de Porfírio.

Figura 1 – Diagrama Biliteral[16]



Fonte: Elaborado pelo autor, baseado em Carroll (1986, p. 79).

Supondo a Classe "Bolos" como Gênero, as linhas do quadrado externo do diagrama acima limitam sua extensão, isto é, todas as coisas que são bolos estão dentro dos limites de tal figura. As Classes "Frescos" e "Não-frescos" são Espécies da Classe "Bolos", cuja divisão dicotômica é representada pela linha horizontal no meio do diagrama, de tal modo que todos os bolos frescos ocupam as células 1 e 2 e todos os não-frescos ocupam as células 3 e 4. As Classes "Bons" e "Não-bons" também são Espécies da Classe "Bolos", com a divisão dicotômica representada pela linha vertical no meio do diagrama, de tal modo que todos os bolos bons ocupam as células 1 e 3 e todos os bolos não-bons ocupam as células 2 e 4. Segue-se que todos os bolos frescos e bons ocupam a célula 1 e que ninguém vai querer comer os bolos da célula 4.

Dada a formação de uma Classe a partir de um certo Gênero, caso existam membros da Classe-Gênero com o Adjunto determinado, servindo como instâncias da Classe-Espécie, tal Classe é chamada de Real; caso não existam membros, isto é, a Classe-Espécie é vazia, então tal Classe é chamada de Irreal ou Imaginária (CARROLL, 1986, p. 60).

A distinção entre Classes Reais e Imaginárias não é análoga à distinção entre Classes cujos membros existem no mundo real e Classes cujos

[16] O diagrama biliteral foi apresentado originalmente em *The Game of Logic* (CARROLL, 1886) e recebeu uma exposição sistemática em *Symbolic Logic* (CARROLL, 1986).

membros não existem no mundo real – do contrário, o País das Maravilhas seria um local muito solitário. Assumir que uma Classe é Imaginária significa apenas a sua vacuidade, isto é, nenhum membro de sua Classe-Gênero pode instanciá-la.

Dada sua caracterização extensional de termos, aceitar que diferentes Classes possam ser Imaginárias é um erro de Carroll na medida em que sua concepção de "Classe" e "membro" é análoga à concepção de "conjunto" e "elemento" em teoria dos conjuntos[17]. Tal analogia é precisa na relação entre a noção carrolliana de Universo do Discurso e a noção conjuntista de conjunto universo (denotado por *U*), mas apresenta problemas na relação entre as noções de Classe Imaginária e conjunto vazio (denotado por Æ).

Dois conjuntos são diferentes apenas se um deles possui ao menos um elemento que não é elemento do outro conjunto. Um conjunto vazio não possui elementos, logo, não pode haver qualquer elemento de um conjunto vazio que não seja elemento de outro conjunto vazio, isto é, existe apenas um conjunto vazio em teoria dos conjuntos[18], tal como deveria haver apenas uma Classe Imaginária sob tal analogia.

A distinção entre intenção (compreensão) e extensão de termos foi introduzida originalmente por Gottfried Leibniz (ABBAGNANO, 2007, p. 217), mas sua discussão contemporânea remonta ao célebre artigo *Sobre o Sentido e a Referência* de Gottlob Frege[19] (2009). A obra de Frege só recebeu atenção internacional em 1903, graças à sua influência no trabalho de Bertrand Russell, isto é, cinco anos após a morte de Carroll e possivelmente desconhecida por ele.

Mesmo que Carroll não estivesse a par da noção contemporânea de intensão e extensão, obviamente conhecia as noções de conotação e denotação, que remontam à escolástica tardia e eram usadas em sua época sob os pa-

[17] A elaboração moderna da teoria dos conjuntos é atribuída a Georg Cantor (1845-1918). Para uma apresentação introdutória, indicamos Mortari (2001, p. 42-60).

[18] A existência de um único conjunto vazio é garantida pelo *Princípio de Extensionalidade*, determinando que se dois conjuntos possuem exatamente os mesmos elementos, então são o mesmo conjunto e não conjuntos diferentes (MORTARI, 2001, p. 45-46).

[19] Publicado originalmente em 1892.

râmetros da definição proposta por John Stuart Mill (ABBAGNANO, 2007, p. 216). Segundo Stuart Mill (1984, p. 98-101):

> Conotativo é o termo que denota um sujeito e implica um atributo. […]. Todos os nomes concretos gerais são conotativos. A palavra homem, por exemplo, denota Pedro, João e um número indefinido de outros indivíduos dos quais, tomados como uma classe, ela é o nome. […]. Das observações precedentes facilmente se concluirá que sempre que os nomes dados aos objetos contenham alguma informação sobre o objeto –isto é, sempre que tenham propriamente alguma significação–, esta significação reside não no que eles denotam, mas no que conotam.

Guardando semelhanças com a noção de conotação de Stuart Mill, Carroll propõe um modelo híbrido: embora caracterize os termos de um ponto de vista extensional, requisito à sua semântica de termos negativos, a existência de múltiplas Classes Imaginárias indica que tal caracterização não é feita sob os parâmetros restritos da noção contemporânea de extensão de conjuntos.

A silogística carrolliana opera sob uma caracterização extensional de termos, mas a individualização de cada Classe não é pautada por sua extensão e sim pela intensão do termo ao qual está associada, possibilitando que Classes Imaginárias coexistam na medida em que sejam associadas a termos com conotações diferentes.

1.2.2 O nonsense associado à lógica

Dada a apresentação das noções lógicas acima, as relações mantidas com o *nonsense* se mostram evidentes, pois o discurso de cada personagem das obras de Carroll só pode ser compreendido se sua interpretação for limitada por um Universo do Discurso respectivo à Classe "Coisas" que constitui o Universo *nonsense* que lhe é suposto, onde coisas que parecem familiares podem ter Adjuntos excêntricos.

Cada coisa que compõe nossa realidade é comumente associável apenas a um certo conjunto de Adjuntos, por mais variado que ele seja, por exemplo, um pedaço de bolo pode ser saboroso ou estar podre, ser nutritivo ou estar envenenado. Embora as coisas que compõem os Universos *nonsense* de

Carroll pareçam familiares às coisas de nossa realidade, sempre podem ter um Adjunto excêntrico aos que costumamos associar a elas, pois um pedaço de bolo, com características similares aos bolos que conhecemos, também pode alterar radicalmente a altura de quem o comer (CARROLL, 2002, p. 18-19).

Não há nenhum problema lógico em assumir que todos os porcos que sofrem de vertigem sejam tratados com respeito (CARROLL, 1986, p. 378-380), mas associar tais características insólitas aos porcos nos deixa perplexos, causando uma espécie de estranheza que só pode ser respondida com a própria aplicação de noções lógicas, não limitadas às características contingentes de nossa realidade, mas usadas com correção para desvendar as maravilhas dos Universos *nonsense* de Carroll.

Segue-se que a noção lógica de "Universo do Discurso" é central ao *nonsense* carrolliano, sendo um requisito básico de sua própria existência na medida em que não se confunde com o absurdo ou a ausência de sentido, mas emerge justamente de sentidos alternativos cuja interpretação só é possível quando limitada por um Universo do Discurso insólito, compreendido como um Gênero de Coisas de um Universo alternativo à nossa realidade, no qual as coisas podem ter Adjuntos excêntricos àqueles que possuem em nosso Universo. "Dito de outro modo, a literatura nonsense tem sentido, mas um sentido diferente" (MONTOITO, 2019, p. 37).

Embora sempre suposta, apenas a noção de "Universo *nonsense*" do Discurso não explica todo o *nonsense* de Carroll, que também se apoia em outras noções lógico-teóricas.

É necessário que cada termo seja compreendido como um Nome que atua como signo convencional para representar uma Espécie que poderia ser representada por qualquer outro signo arbitrário. Por exemplo, Alice só consegue entender Humpty Dympty quando ele esclarece que escolhe de modo idiossincrático os Nomes que usa para representar as Espécies de Coisas do Universo do Discurso suposto à sua fala: "'Quando eu uso uma palavra', disse Humpty Dumpty num tom bastante desdenhoso, 'ela significa exatamente o que quero que signifique: nem mais nem menos'" (CARROLL, 2002, p. 204).

Embora a relação entre um Nome e a Classe de Coisas que ele representa possa ser convencionada de modo arbitrário para cada termo, a intensão do falante ao usar um Nome determina a individualização da Classe representada, de tal modo que não há apenas uma Classe Imaginária (vazia), mas Classes Imaginárias distintas podem coexistir se representadas por Nomes usados com intensões diferentes. De fato, Classes Imaginárias não apenas existem nos Universos *nonsense* de Carroll, mas talvez possam até ser vistas:

> "[...]. Dê uma olhada na estrada, e diga-me se pode ver algum deles."
> "Ninguém à vista", disse Alice,
> "Só queria ter olhos como esses", observou o Rei num tom irritado. "Ser capaz de ver Ninguém! E à distância! Ora, o máximo que eu consigo é ver pessoas reais, com esta luz!" (CARROLL, 2002, p. 214)

Supondo "pessoas" como Universo do Discurso, o Rei se surpreende ao compreender a fala de Alice como um relato de sua percepção da Classe Imaginária "pessoas na estrada", afirmando que só consegue ver pessoas reais, isto é, só consegue perceber a Classe "pessoas na estrada" se ela for uma Classe Real.

Dado que "estar na estrada" não é um Adjunto excêntrico aos membros da Classe "pessoas" de nosso Universo, o fragmento supracitado também demonstra que nem todo *nonsense* carrolliano surge da fórmula "coisa com Adjunto insólito", mas pode emergir de qualquer manipulação criativa de noções lógico-teóricas.

Por exemplo, embora Carroll defenda que nenhum Adjunto pode existir sem uma coisa a qual pertença (1886, p. 2), curiosamente, a possibilidade de tal existência é o mote de um dos fragmentos mais célebres da literatura carrolliana – a efêmera existência do sorriso como um Adjunto sem o Gato de Cheshire ao qual pertence: Alice pede ao Gato de Cheshire que pare de aparecer e sumir repentinamente, pois está lhe causando vertigens.

> "Está bem", disse o Gato; e dessa vez desapareceu bem devagar, começando pela ponta da cauda e terminando com o sorriso, que persistiu algum tempo depois que o resto de si fora embora.

> "Bem! Já vi muitas vezes um gato sem sorriso", pensou Alice; "mas um sorriso sem gato! É a coisa mais curiosa que já vi na minha vida!" (CARROLL, 2002, p. 64-65)

Alice considera um sorriso sem gato como a coisa mais curiosa que já viu na vida, mesmo depois de todas as loucuras que a levaram até o final do sexto capítulo de suas aventuras no País das Maravilhas, onde tal evento ocorre, demonstrando o quanto Carroll é avesso à existência isolada de um Adjunto.

O uso da noção de termos negativos como fundamento ao *nonsense* também precisa ser citado, dado que o método diagramático para resolução de silogismos com termos negativos constitui a maior contribuição carrolliana ao desenvolvimento da teoria silogística.

> "Quero dizer, o que é um presente de desaniversário?"
> "Um presente dado quando não é seu aniversário, é claro."
> Alice refletiu um pouco. "Gosto mais de presentes de aniversário", declarou finalmente.
> "Não sabe do que está falando!" exclamou Humpty Dumpty. (CARROLL, 2002, p. 203-204)

Supondo o Gênero "dia comemorativo" como Universo do Discurso, o Nome "aniversário" compreende uma Espécie com o Adjunto "do nascimento de alguém". No fragmento acima, o neologismo "desaniversário" é usado como sinônimo do termo negativo "não-aniversário", isto é, respectivo aos dias comemorativos do não-nascimento de alguém, uma celebração insólita que levaria todos à falência se realmente fosse comemorada.

1.3 O Nonsense Como Instrumento de Popularização da Lógica

O *nonsense* é tradicionalmente categorizado como um subgênero de humor da literatura infantil, mas "Os mais importantes estudiosos do nonsense, com destaque para o alemão Klein Reichert, têm como pressuposto que o nonsense não é um subgênero de humor [...]" (ÁVILA, 1996, p. 17 e 133), inclusive o próprio status de infantil da literatura *nonsense* carrolliana é questionado por Montoito (2019, p. 53), devido à profundidade e às muitas camadas de interpretação de suas obras, e Bertrand Russell, que a considera útil para

adultos "na necessidade de instrução metafísica e saneamento lógico" (RUSSELL, 1996, p. 524, tradução nossa).

Mesmo não sendo um subgênero de humor, o *nonsense* carrolliano é repleto de situações cômicas que são fruto da relação entre dois elementos sempre presentes em suas obras: noções lógicas e um universo de fantasia insólito. A presença basilar de um universo de fantasia também gera outro equívoco: faz com que *Alice no País das Maravilhas* (CARROLL, 2002) seja apresentada como um conto de fadas. Segundo Tolkien (2017, p. 10):

> A definição de conto de fadas – o que é, ou o que deveria ser – não depende, portanto, de nenhuma definição ou relato histórico de elfos ou fadas, mas sim da natureza do Reino Encantado, do próprio Reino Perigoso, e do ar que sopra nessa terra. Não tentarei defini-lo nem descrevê-lo diretamente. […]. Por ora só direi isto: um "conto de fadas" é aquele que toca ou usa o Reino Encantado, qualquer que seja o seu propósito principal, […].

Embora o País das Maravilhas pareça ser um reino de fantasia associável ao conceito de Reino Encantado tolkieniano, a natureza do Reino Encantado exige que ele seja apresentado como "verdadeiro", isto é, suas "maravilhas" não podem ser explicadas por um mecanismo que as considere como uma ficção ou delírio (TOLKIEN, 2017, p. 14). Dado o enquadramento de sonho[20] na obra de Carroll, segue-se que ela não é um conto de fadas: "o elemento de 'sonho' não é mera trama de introdução e conclusão, mas inerente à ação e às transições" (TOLKIEN, 2017, p. 71).

O enquadramento de sonho não se limita às aventuras de Alice, "em quase todas as obras carrollianas, [os sonhos] são cenários narrativos" (MONTOITO, 2019, p. 28), gerando outro equívoco: a associação do *nonsense* carrolliano com o surrealismo, que "tem como pilar a expressão espontânea e automática do pensamento, ditada pelo inconsciente, numa valorização dos aspectos do sonho" (MONTOITO, 2019, p. 45).

Apesar de também valorizar o sonho, o *nonsense* de Carroll pode ser interpretado sob o seu próprio sistema de referências e é repleto de desafios lógicos cuidadosamente elaborados, evidenciando um processo criativo com-

[20] Alice acorda no final do livro: "'Acorde, Alice querida!' disse sua irmã. […]. 'Ah, tive um sonho tão curioso!' disse Alice [...]" (CARROLL, 2002, p. 122).

pletamente distinto dos surrealistas. Diferente deles, Carroll usa o sonho apenas como um recurso narrativo para introduzir os universos de fantasia que possibilitam o surgimento de seu *nonsense*: "[...], o sonho, pelas (im)possibilidades que permite às personagens, é a porta que separa os universos reais e ficcionais [...]" (MONTOITO, 2019, p. 28).

Distante do Reino Encantado dos contos de fada e da expressão do inconsciente dos surrealistas, os universos de fantasia carrollianos são um recurso narrativo para a apresentação de suas surpreendentes manipulações de noções lógico-teóricas.

Os seus cenários de fantasia estão tão relacionados às noções lógicas que compõem a trama que se passará neles que a própria noção de "universo de fantasia carrolliano" é melhor definida sob suas próprias noções lógicas, a saber, universos nos quais as Coisas podem ter Adjuntos excêntricos de tal modo que o Universo do Discurso de cada personagem só pode ser compreendido se determinado pela Classe "Coisas" que lhe é específica, sendo distinta da Classe "Coisas" associada à nossa realidade.

O "Universo *nonsense*" do Discurso é suposto não apenas ao *nonsense* literário, como também ao *nonsense* encontrado nos exemplos e exercícios dos trabalhos mais sérios de Carroll sobre lógica, mas o seu *nonsense* também emerge das mais variadas e criativas manipulações de noções lógico-teóricas, explorando as possibilidades férteis de seus universos de fantasia para, segundo Montoito (2019, p. 27), comunicar noções lógicas.

Embora Carroll seja "apontado por vários autores como um escritor que colocou em xeque a literatura pedagógica de fundo moral [...]" (GARNICA, 2014, p. 17)[21], sua preocupação pedagógica com a comunicação de noções lógicas não pode ser ignorada.

Carroll foi um professor engajado na criação de metodologias alternativas para o ensino de lógica, como demonstra a publicação original de seu método diagramático para a resolução de silogismos (1886), apresentado como um jogo de peças, visando o seu ensino lúdico. Também foi um grande

[21] Os livros infantis da época buscavam ensinar a moral vitoriana com ênfase às punições, o rompimento de Carroll com tal tradição pode ser percebido "claramente nos livros de Alice, nos quais a protagonista não se dobra à ordem vigente (principalmente no que diz respeito às imposições da Rainha) e não recebe punição alguma por isso" (GARNICA, 2014, p. 17-18).

crítico do sistema educacional de sua época e é considerado o principal responsável pela divulgação da lógica durante o século XIX (THE JOY OF LOGIC, 2013), não apenas por seus trabalhos acadêmicos, mas também pela capacidade de desenvolver o raciocínio lógico dos leitores de suas obras literárias.

Segunda Ávila (1996, p. 203):

> A poesia nonsense tem uma especificidade que é responsável pela atração permanente que exerce sobre leitores e teóricos, desde sua criação. Sua especificidade reside em algo que deixa o leitor suspenso entre o riso e a perplexidade, entre a estranheza e a identificação, como se aquilo ao mesmo tempo lhe dissesse respeito e não dissesse respeito a coisa nenhuma.

Generalizando a tese de Ávila sobre poesia *nonsense*, segue-se que o *nonsense* de Lewis Carroll pode ser compreendido como a forma pela qual ele deixa os seus leitores "suspensos entre o riso e a perplexidade" por meio das peculiares e inovadoras manipulações de noções lógico-teóricas que apresenta graças às possibilidades oriundas de seus universos ficcionais: seja extraindo uma conclusão valida e insólita de um silogismo cujos membros do Universo do Discurso possuem Adjuntos exóticos, fomentando o raciocínio lógico, seja apresentando propriedades de noções teóricas complexas em meio à trama de suas obras literárias graças às maravilhas de seus universos de fantasia. Em ambos os casos, o *nonsense* é usado como um meio para induzir instrução lógica nos seus leitores.

2. O Nonsense Wittgensteiniano

A análise da noção de *nonsense* já foi objeto do trabalho de muitos filósofos ilustres, como Husserl (1900), que distinguiu expressões cujo *nonsense* advém da forma (por exemplo: "ou verde amanhã então") de expressões cujo *nonsense* advém do conteúdo (por exemplo: "a mesa é sábia"), e Deleuze, que dedicou atenção especial ao *nonsense* carrolliano em *Lógica do Sentido* (1974).

Para Deleuze (1974), apenas proposições com sentido são verdadeiras ou falsas, de modo que proposições *nonsense* não possuem valor de verdade,

mas o *nonsense* carrolliano é categorizado como um tipo especial, compreendido não como um simples fruto da ausência de sentido e sim como o resultado de um trabalho sobre as condições de possibilidade de sua produção, florescendo em um solo semelhante ao ocupado por crianças que buscam se expressar balbuciando sons sem significado: "Se Lógica do Sentido [de Deleuze] é em parte uma leitura do trabalho de Carroll, é porque ninguém conhecia as condições para a produção do sentido melhor do que Carroll" (SMITH; PROTEVI, 2018, tradução nossa).

Entre todas as abordagens filosóficas à noção de *nonsense*, sem dúvida a mais popular e influente foi feita por Wittgenstein, apresentada originalmente na obra *Tractatus Logico-Philosophicus* (2001) e rearticulada em *Investigações Filosóficas* (1975).

Examinar a noção de *nonsense* wittgensteiniana exige uma breve apresentação de seus pressupostos. O *Tractatus* (2001, p. 135-165) apresenta uma concepção segundo a qual o mundo consiste na existência de estados de coisas e toda proposição com sentido consiste na figuração lógica destes estados de coisas, isto é, tal como uma pintura realista representa uma paisagem na medida em que seus elementos gráficos seguem uma proporção e um ordenamento adequado às coisas que compõe a paisagem, toda proposição com sentido é a figuração de um estado de coisas sob "a possibilidade de que as coisas estejam uma para as outras tal como os elementos da figuração" (WITTGENSTEIN, 2001, p. 143).

Segundo Wittgenstein, a proposição "mostra seu sentido. A proposição mostra como estão as coisas se for verdadeira e diz que estão assim" (2001, p. 169). Ou seja, a proposição com sentido mostra o estado de coisas que lhe é correspondente, mas é o estado de coisas que define o valor de verdade da proposição que pretende figurá-lo: a proposição será verdadeira se for uma figuração adequada e falsa se não for.

Segue-se que toda proposição com sentido é contingente, de tal modo que tautologias e contradições são proposições sem sentido (*sinnlos*). Por exemplo, a proposição tautológica "Todo quadrado é não-redondo", que não

mostra um estado de coisas, expressando apenas o que já é logicamente suposto aos seus elementos.

> A proposição mostra o que diz; a tautologia e a contradição, que não dizem nada. A tautologia não tem condições de verdade, pois é verdadeira incondicionalmente; e a contradição, sob nenhuma condição. Tautologia e contradição não têm sentido. (WITTGENSTEIN, 2001, p. 197)

Dada a apresentação das noções de proposição com e sem sentido, segue-se a abordagem wittgensteiniana de *nonsense* (*unsinn*)[22]: enunciados destituídos de valor veritativo.

Enunciados *nonsense* não são figurações de estados de coisas que balizam as suas condições de verdade, não sendo verdadeiros ou falsos devido às contingências factuais e não se caracterizando como proposições com sentido, mas também não são necessariamente verdadeiros ou necessariamente falsos em virtude da relação lógica entre seus próprios elementos, como tautologias e contradições, também não se caracterizando como proposições sem sentido.

Uma vez que todo estado de coisas passível de figuração é um fato que está "dentro" do mundo, o que está "fora" do mundo não pode ser representado pela linguagem. Os enunciados da metafísica tradicional, da ética e da estética, embora não sejam tautológicos ou contraditórios, são pretensas predicações de conceitos formais, enunciados *nonsense* que tentam capturar o mundo como um todo, sem figurarem nada "dentro" do mundo que possa balizar suas condições de verdade.

Para Wittgenstein, a filosofia "não é uma teoria", mas uma atividade cujo fim "é o esclarecimento lógico dos pensamentos", de tal modo que "o resultado da filosofia não são 'proposições filosóficas', mas sim tornar proposições claras" (2001, p. 177). Segue-se que a noção de *nonsense* wittgensteiniana incide sobre todos os problemas filosóficos, mostrando "que a formulação desses problemas repousa sobre o mal entendimento da lógica de nossa linguagem" (WITTGENSTEIN, 2001, p. 131).

A concepção de linguagem expressa no *Tractatus* (2001) é revista por seu autor em *Investigações Filosóficas* (WITTGENSTEIN, 1975), afirmando

[22] Luiz Henrique Lopes dos Santos (WITTGENSTEIN, 2001) traduz "unsinn" como "contrassenso".

que não pode haver uma concepção geral de linguagem, pois ela não compõe um todo homogêneo, apresentando sua nova abordagem em analogia com a noção de jogo: futebol e xadrez são corretamente classificados como jogos – mesmo com tão pouco em comum, assim como existem diferentes jogos de linguagem que partilham entre si apenas certas semelhanças de família, tal como familiares compartilham traços genéticos comuns mesmo não havendo uma característica presente em todos os membros da família.

> A tese dessa obra é que toda linguagem é uma espécie de jogo que segue determinadas regras, e todos os jogos linguísticos têm o mesmo valor. Portando, a única regra para a interpretação de um desses jogos é, segundo Wittgenstein, o uso que dela se faz; e como a filosofia não tem outra tarefa senão a análise lógica da linguagem, o esclarecimento das expressões linguísticas em seu uso corrente é tarefa própria da filosofia. (ABBAGNANO, 2007, p. 719)

O Wittgenstein das *Investigações* (1975) abandona a posição segundo a qual todo *nonsense* pode ser reconhecido com uma análise lógica das proposições, tal como expresso no *Tractatus* (2001), dado que nem todo jogo de linguagem se submete às mesmas regras. Ainda assim, enunciados *nonsense* continuam sendo compreendidos como destituídos de valor veritativo, pois são frutos de distintos erros e confusões que os tornam inadequados aos parâmetros estabelecidos pelas regras do jogo de linguagem ao qual pertencem.

Embora não defenda um método filosófico singular para identificar o *nonsense*, Wittgenstein (1975, p. 62) propõe que a atividade filosófica pode ser realiza por diferentes métodos.

> Um dos métodos que Wittgenstein apresenta nas Investigações Filosóficas se aproxima, justamente, daquilo que Carroll faz em seus poemas e romances. Quando a estrutura de uma proposição não nos permite perceber que se trata de um nonsense, o filósofo da linguagem, esse terapeuta das palavras, pode imaginar como seria um mundo a partir dessa proposição. Nesse caso, o caráter nonsense da proposição ficaria claro com a ficção. No campo da pura imaginação, conheceríamos as consequências absurdas de determinadas proposições [...]. Com a descrição desse mundo fictício, o filósofo terapeuta pode compará-lo, em bloco, ao mundo em que vivemos, como quem compara o sonho e a vigília. (SILVEIRA, 2010)

Embora as duas obras de Wittgenstein apresentem concepções distintas de *nonsense*, duas ideias centrais permanecem inalteradas: enunciados

nonsense não possuem valor de verdade e a tarefa do filósofo é evidenciar o *nonsense*.

3. A Relação Entre o Nonsense de Carroll e Wittgenstein

Carroll apresenta uma redefinição da noção de silogismo devido ao seu acréscimo de termos negativos à teoria silogística. Segundo Carroll, dado um trio de proposições, todo silogismo válido deve cumprir três cláusulas (1986, p. 107, tradução nossa):

> (1) Todos os seis Termos são Espécies do mesmo Gênero.
> (2) Qualquer dupla [de proposições] sempre vai conter entre si um par de Classes
> codivisionais.
> (3) As três Proposições estão relacionadas de tal maneira que, se as duas primeiras
> fossem verdadeiras, a terceira seria verdadeira.

Segundo Sautter (2015, p. 13), a primeira cláusula é ontológica, pois define que todos os termos pertencem ao mesmo Universo do Discurso; a segunda é epistemológica, pois possibilita que informações sobre a relação mantida entre dois termos seja obtida com base na relação que eles mantém com um terceiro termo; e a terceira cláusula é exclusivamente lógica, pois determina que a terceira proposição seja o resultado de uma inferência válida das outras duas.

A terceira cláusula não exige que as proposições de um silogismo válido sejam verdadeiras, mas exige que a hipótese de suas verdades seja possível, isto é, devem ser proposições cujo valor veritativo pode ser o verdadeiro. Segue-se que o *nonsense* usado por Carroll nos exemplos de silogismo apresentados em seus trabalhos acadêmicos não é, *stricto sensu*, *nonsense* no sentido wittgensteiniano, pois é constituído de proposições com sentido, que podem ser verdadeiras.

Todos os exemplos de silogismo com apelo *nonsense* usados por Carroll em seus trabalhos acadêmicos nascem da fórmula "coisa com Adjunto insólito", mas ela não esgota o *nonsense* de suas obras literárias, que emerge das mais distintas manipulações de noções lógicas, constituindo uma audaciosa

exploração dos limites do uso da linguagem e amostras do resultado da subversão de suas regras.

> Autores vitorianos de nonsense, como Lewis Carroll, apontaram problemas que deveriam manter os filósofos da linguagem ocupados pela maior parte do século XX. Os livros de Alice despertam a consciência para as armadilhas da linguagem (escrita), armadilhas que em nossa vida cotidiana quase nunca produzem problemas, mas que estão no coração do discurso filosófico. (WAGNER, 2012, p. 215, tradução nossa)

Cabe ressaltar que Wittgenstein cita Carroll ao expressar suas considerações sobre a capacidade das palavras de designar objetos:

> Quando dizemos: "cada palavra da linguagem designa algo", com isso ainda não é dito absolutamente nada; a menos que esclareçamos exatamente qual a diferença que desejamos fazer. (Pode bem ser que queiramos diferenciar as palavras da linguagem de palavras 'sem significação', como ocorrem nas novelas de Lewis Carroll, ou de palavras como "la-la-ri-la-la" numa canção.) (WITTGENSTEIN, 1975, p. 18)

Wittgenstein atribuí "sem significação" e não *nonsense* às palavras que ocorrem nas novelas de Carroll. O termo "sem significação" foi empregado para traduzir "*ohne Bedeutung*" e não "*unsinn*" ou "*sinnlos*", usados respectivamente para traduzir "*nonsense*"[23] e "sem sentido".

Embora Wittgenstein possa ter usado o termo "sem significação" em vez de "*nonsense*" por referir-se às palavras e não às "proposições" de Carroll[24], ele o usa entre aspas, sugerindo que talvez compreenda o *nonsense* carrolliano de um modo especial, como também sugere a seguinte citação (WITTGENSTEIN, 1975, p. 106):

> "Mas, em contos de fada, um pote pode também ver e ouvir!" (Certo; mas ele pode também falar.)
> "Mas um conto apenas inventa o que não ocorre; não diz nenhum absurdo." – Isto não é tão simples. É inverdade ou absurdo dizer que um pote fala? Fazemos uma imagem clara das condições nas quais se diria que um pote fala? (Mesmo um poema-absurdo não é um absurdo no mesmo sentido que o balbuciar de uma criança.)

[23] José Carlos Bruni usa "absurdo" (WITTGENSTEIN, 1975) e João José R. L. de Almeida usa "contrassenso" (WITTGENSTEIN, [s./d.]) para traduzir "*unsinn*".

[24] Dado que apenas proposições podem ter valor de verdade e apenas enunciados analisados em um contexto de análise proposicional podem ser classificados como *nonsense*, isto é, destituídos de valor veritativo. Do contrário, todas as palavras analisadas isoladamente são *nonsense*, pois não são proposições e, portanto, são destituídas de valor veritativo.

> Sim; dizemos de seres inanimados que eles têm dor: brincando com bonecas, por exemplo. Mas este emprego é secundário. Imaginemos o caso em que se dissesse apenas de seres inanimados que eles têm dor; e que nos compadecêssemos apenas de bonecas!

No fragmento supracitado, Wittgenstein busca esclarecer a distinção entre proposições falsas e *nonsense* – enquanto exemplifica como experimentos mentais podem ser usadas como um método filosófico para evidenciar o *nonsense*. Mas Wittgenstein também expressa que "um poema-nonsense não é um nonsense no mesmo sentido que o balbuciar de uma criança", novamente parecendo compreender a literatura *nonsense* carrolliana de um modo especial.

Segundo Pitcher (1965, p. 592, tradução nossa), "Carroll exerceu uma profunda influência no [trabalho] tardio de Wittgenstein":

> O nonsense de Wittgenstein e Carroll produzem uma perplexidade extrema: Alice fica constantemente perplexa e confusa com o nonsense que ouve no decorrer de suas aventuras, assim como os filósofos, segundo Wittgenstein, ficam perplexos e confusos com o nonsense que eles mesmos inconscientemente expressam. (PITCHER, 1965, p. 611, tradução nossa)

Mesmo convicto da influência de Carroll no trabalho tardio de Wittgenstein, Pitcher sustenta que suas atitudes em relação ao *nonsense* divergiam extremamente. Segundo Pitcher (1965, p. 611, tradução nossa):

> […], Wittgenstein e Carroll tinham atitudes radicalmente diferentes em relação ao nonsense: ele torturava Wittgenstein e deliciava Carroll. Carroll virou as costas para a realidade a partir do mundo (maravilhoso) do mito e da fantasia. Wittgenstein, sendo um filósofo, empenhou todos os seus esforços para nos arrastar de volta à realidade distinta do mundo (horrível) do mito e da fantasia. [...] a atitude, certamente, é fundamentalmente diferente.

Pitcher justifica sua posição alegando que "[…] as mesmas confusões com as quais Wittgenstein cobra os filósofos foram empregadas deliberadamente por Carroll para efeito cômico" (PITCHER, 1965, p. 593. tradução nossa), de tal modo que Wittgenstein considera o *nonsense* como "perigosamente apto para enganar o filósofo", enquanto "Carroll, por outro lado, simplesmente se diverte" (PITCHER, 1965, p. 605, tradução nossa).

Como já analisado (ver subseção 1.3), o *nonsense* carrolliano não serve apenas para a diversão e não pode ser reduzido a um subgênero de humor, pois se trata de uma estrutura narrativa que possibilita a manipulação de noções lógico-teóricas de forma inovadora, constituindo um meio lúdico usado para induzir instrução e popularizar a lógica. Inclusive o clássico *Alice* (CARROLL, 2002) foi criado por Carroll buscando levar um módico de instrução lógica às irmãs Liddell[25] enquanto as distraia durante um passeio de barco no qual estavam aos seus cuidados (GARDNER, 2002c, p. 7-9).

Também é errado afirmar que Carroll "virou as costas para a realidade" em suas obras *nonsense*, como afirma Pitcher (1965, p. 611). Carroll não escreveu contos de fada que supõem a realidade de seus universos *nonsense* (ver subseção 1.3), pois a trama *nonsense* é sempre cuidadosamente estruturada sob a moldura de um sonho e objetiva induzir instrução lógica. Embora Carroll utilizasse universos de fantasia para induzir destrezas lógicas aos seus leitores, ele defendia que a utilidade dessas destrezas se aplicava ao mundo real, principalmente à detecção de falácias:

> Você acredita que o principal uso da Lógica, na vida real, é o de deduzir conclusões viáveis, e o de convencer-se que as conclusões, deduzidas por outras pessoas, estão corretas? Quisera fosse assim! A sociedade estaria muito menos sujeita ao pânico e a outras ilusões, e a vida política, sobretudo, seria algo totalmente diferente, caso a maioria dos argumentos transmitidos e espalhados pelo mundo fossem corretos! Mas, receio, é exatamente o contrário. Para cada par de premissas viáveis (aquelas que levam a uma conclusão lógica) que você encontra ao ler um jornal ou uma revista, você provavelmente encontrará cinco que não levam a conclusão alguma, e mesmo quando as premissas são viáveis, para cada instância na qual o escritor extrai uma conclusão correta, há provavelmente dez nas quais o escritor extrai uma conclusão incorreta […]. O uso principal que você fará dessa destreza lógica, […], será o de detectar 'Falácias' […]. (CARROLL, 1887, p. 32-33)

Ao acordar do sonho *nonsense*, o protagonista das obras de Carroll está pronto para aplicar em vigília as mesmas noções lógicas que moldaram o seu sonho, tal como o leitor que, ao fechar o livro, pode aplicar o módico de instrução lógica que recebeu durante sua leitura à própria realidade e detectar

[25] Filhas de Henry Liddell, amigo de Carroll e deão do Christ Church, onde Carroll lecionava. Entre elas estava Alice Pleasance Liddell (1852-1934), que serviu de inspiração para a protagonista homônima.

as falácias que o cercam. Sautter (2015, p. 9) destaca que a detecção de falácias "não é um mero capítulo da lógica carrolliana, mas é o próprio coração dela".

Segue-se que o *nonsense* carrolliano não vira as costas para a realidade, pois se direciona justamente a ela enquanto instrumento didático para a popularização de destrezas lógicas – tão aplicáveis à realidade que, se popularizadas, podem tornar a sociedade menos sujeita ao pânico e a outras ilusões.

Compreendendo o *nonsense* carrolliano como o resultado das peculiares e inovadoras manipulações de noções lógico-teóricas que Carroll buscou apresentar aos seus leitores, segue-se que devemos compreender o seu uso como análogo ao uso de enunciados *nonsense* que Wittgenstein faz em suas próprias obras:

> Não há dificuldade alguma de se levar às últimas consequências o que vem dito no Prefácio do Tractatus, ou na penúltima proposição do livro. Nenhuma proposição do Tractatus faz o menor sentido – nem a primeira, nem a última, […]. No entanto, são capazes de produzir efeitos e, por pertencerem a um sistema linguístico usado por seres humanos, podem produzir efeitos mais ou menos previsíveis nesses seres. Podem levar esses seres a assumirem determinadas atitudes, fazerem ou deixarem de fazer determinadas coisas, falarem ou deixarem de falar outras tantas.
> […] O Tractatus não pretende ser um discurso verdadeiro, mas pretende, sim, induzir seu leitor a uma perspectiva que o autor do livro pretende que seja a perspectiva correta sobre a linguagem. (CUTER, 2015, p. 36)

Wittgenstein usa o *nonsense* para induzir uma certa perspectiva sobre a linguagem, buscando exorcizar a filosofia do *nonsense* fruto das armadilhas da linguagem. Carroll não propõe uma filosofia com enunciados *nonsense*, apenas usa o *nonsense* com a finalidade pedagógica de comunicar noções lógico-teóricas, explorando as armadilhas da linguagem como brincadeiras que lhe permitem conceber os cenários hipotéticos adequados para testar e explorar consequências lógicas.

Nas palavras de Pitcher (1965, p. 611, tradução nossa): "O mesmo terreno lógico que é um playground para Carroll, é um campo de batalha para Wittgenstein". Mas, diferente da posição de Pitcher (1965), Carroll e Wittgenstein não tinham atitudes radicalmente diferentes em relação ao *nonsense*,

pois ambos o usaram para produzir efeitos instrutivos em seus leitores – embora o propósito de Carroll fosse a instrução lógica e o de Wittgenstein a instrução sobre uma perspectiva que ele julgava correta sobre a linguagem.

Considerações Finais

Dado o desejo de divulgar e popularizar a lógica nutrido por Carroll, conclui-se que o seu *nonsense* deve ser compreendido como um meio para esse fim. Todos os diálogos *nonsense* de suas obras supõem um Universo do Discurso adequado à sua compreensão, mas o seu *nonsense* não é limitado por uma única noção teórica, nascendo das mais variadas e criativas manipulações de noções lógico-teóricas.

Embora nem todo enunciado *nonsense* carrolliano seja associável à concepção de *nonsense* wittgensteiniana em um sentido restrito, dado que alguns de seus enunciados são apenas falsos e não destituídos de valor veritativo, a maioria é completamente congruente com a noção de Wittgenstein e, especialmente em suas obras literárias, o seu *nonsense* parece fruto justamente da aplicação de um dos métodos que Wittgenstein propõe para identificar o *nonsense*, tal como sugere Silveira (2010).

Além disso, "[…] sondar os limites da linguagem era o empreendimento de ambos, Carroll e Wittgenstein" (WAGNER, 2012, p. 215, tradução nossa), e o *nonsense* presente na obra de ambos emerge como os resultados de tal empreendimento, compartilhando objetivos afins.

Diferente de Pitcher (1965), conclui-se que Carroll e Wittgenstein tinham atitudes análogas em relação ao *nonsense*, a saber, ambos usaram o *nonsense* como um meio para induzir ideias instrutivas nos seus leitores. Carroll usou o *nonsense* para popularizar noções lógicas, enquanto Wittgenstein usou o *nonsense* para apresentar a perspectiva que julgava correta sobre a linguagem.

Referências

ABBAGNANO, N. *Dicionário de Filosofia.* 5 Ed. Tradução original coordenada por Alfredo Bossi; revisada e estendida por Ivone Castilho Benedetti. São Paulo: Martins Fontes, 2007. 1210 p.

ABELES, F. F. “Lewis Carroll's visual logic”. *History and Philosophy of Logic*, v. 28, n. 01, p. 1-17, 2007. p. 1-17.

ANGELELLI, I. *Studies on The Gottlob Frege and Traditional Philosophy.* Dordrecht, Holanda: D. Reidel Publishing Company, 1967. p. 12-13.

ARISTÓTELES. “Periérmeneias”. In: ARISTÓTELES. *Organon*. Tradução: Pinharanda Gomes. Lisboa: Guimarães Editores, 1985. p. 119-173.

ÁVILA, M. *Rima e Solução: a Poesia Nonsense de Lewis Carroll e Edward Lear*. São Paulo: Annablume, 1996. 234 p.

BALAGUER, M. “Platonism in Metphysics”. In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy.* Stanford, California. Spring, 2016. Disponível em: https://plato.stanford.edu/archives/spr2016/entries/platonism/. Acesso em: 25 jan. 2020.

BOOLE, G. *An Investigation of the Laws of Thought: on which are founded: The Mathematical Theories of Logic and Probabilities*. Cambridge: McMillan and Co., 1854. 425 p.

CARROLL, L. *Through The Looking-Glass: And What Alice Found There*. London: MacMillan and Co., 1872. p. 21-22.

CARROLL, L. *The Game of Logic*. London: MacMillan and Co., 1886. 124 p.

CARROLL, L. *The Game of Logic*. 2. ed. London: MacMillan and Co., 1887. Tradução para o português em elaboração: CARROLL, L. *O jogo da lógica*. Tradução: Frank Thomas Sautter. London: College Publications.

CARROLL, L. “What the Tortoise Said to Achilles”. *Mind*, v. 14, Abril, 1895. p. 278-280.

CARROLL, L. *Symbolic Logic: Lewis Carroll's*. 6. ed. Rev., ampl. e atual. New York: Clarkson Potter, 1986. 514 p.

CARROLL, L. *Alice: Edição comentada*. Tradução: Maria Luiza X. de A. Borges. Rio de Janeiro: Jorge Zahar Ed., 2002. 303 p.

CARROLL, L. *The Complete Stories and Poems of Lewis Carroll*. New Lanark: Geddes & Grosset, 2005. p. 148.

CARROLL, L. *The Pamphlets of Lewis Carroll*. Vol. 4. Compilado por Francine F. Abeles. New York: University Press of Virginia, 2010. 271 p.

CARROLL, L. *Jaguadarte*. Tradução: Augusto de Campos. São Paulo: Editora Nhambiquara, 2014. 56 p.

COHEN, M. N. *Lewis Carroll: Uma biografia*. Tradução: Raffaella de Filippis. Rio de Janeiro: Record, 1998. 669 p.

CUTER, J. V. G. "Como entender um contrassenso e reconhecer sua correção". *Dissertatio*. Volume Suplementar. Dossiê Wittgenstein. Junho, 2015. p. 31-43.

DELEUZE, G. *Lógica do Sentido*. Tradução: Luiz Roberto Salinas Fortes. São Paulo: Perspectiva, 1974. 342 p.

DODGSON, C. L. *Doublets: A Word-Puzzle, by Lewis Carroll*. 2 Ed. London: MacMillan, 1880.

FREGE, G. *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle, Alemanha: Louis Nebert, 1879. 89 p.

FREGE, G. "Sobre o Sentido e a Referência". In: FREGE, G. *Lógica e Filosofia da Linguagem.* 2 Ed. Tradução: Paulo Alcoforado. São Paulo: Editora da Universidade de São Paulo, 2009. p. 129-158.

GARDNER, M. "Introdução à 1ª Edição". In: CARROLL, L. *Alice: Edição comentada*. Tradução: Maria Luiza X. de A. Borges. Rio de Janeiro: Jorge Zahar Ed., 2002a. p. vii-xxi.

GARDNER, M. "Introdução à 2ª Edição". In: CARROLL, L. *Alice: Edição comentada*. Tradução: Maria Luiza X. de A. Borges. Rio de Janeiro: Jorge Zahar Ed., 2002b. p. xvii-xxi.

GARDNER, M. "Notas". In: CARROLL, L. *Alice: Edição comentada*. Tradução: Maria Luiza X. de A. Borges. Rio de Janeiro: Jorge Zahar Ed., 2002c. p. 7-302.

GARNICA, A. V. M. "Euclides, aliado de Alice, e seus rivais oitocentistas". In: CARROLL, L. *Euclides e seus rivais modernos*. Tradução: Rafael Montoito. São Paulo: Editora Livraria da Física, 2014. p. 9-21.

HUSSERL, E. *Logische Untersuchungen*. Leipzig: Verlag Von Veit & Comp., 1900. 257 p.

KENNY, A. *História Concisa da Filosofia Ocidental*. Tradução: Desidério Murcho, Fernando Martinho, Maria José Figueiredo, Pedro Santos e Rui Cabral. Lisboa: Temas e Debates — Actividades Editoriais, 1999. p. 438.

KEYNES, J. N. *Studies and Exercises in Formal Logic*. 4. Ed. London: Macmillan and Co., 1906. 548 p.

LEAR, E. *A Book of Nonsense*. London: Thomas McLean, 1846. 112 p.

LINDEMANN, J. L. *A Lógica de Lewis Carroll.* Dissertação (Mestrado em Filosofia) – Programa de pós-graduação em Filosofia. Universidade Federal de Santa Maria, Santa Maria, 2017. 121 p.

LINDEMANN, J. L. "Árvores de Carroll-Smullyan". Artigo submetido: *Dissertatio*, 2020. 26 p.

LOUX, M. J. *Metaphysics: A Contemporary Introduction*. London and New York: Routledge, 1998.

MALCOLM, N. *The origins of English nonsense*. London: Harper Collins, 1997. p. 14.

MILLER, D. "Word Games for Formal Logic". In: *FIRST INTERNATIONAL CONGRESS ON TOOLS FOR TEACHING LOGIC*. Salamanca: ARACNE, 2000. p. 1-4. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.127.5330&rep=rep1&type=pdf. Acesso em: 19 jun. 2020.

MILL, J. S. "Sistema de Lógica Dedutiva e Indutiva: Exposição dos Princípios da Prova e dos Métodos de Investigação Científica (Seleção)". In: MILL, J. S. *Sistema de lógica dedutiva e indutiva e outros textos.* 3 Ed. Tradução: João Marcos Coelho e Pablo Rubén Mariconda. São Paulo: Abril Cultural, 1984. p. 79-255. (Os pensadores)

MONTOITO, R. *Lógica e Nonsense nas Obras de Lewis Carroll: Silogismos e Tontogismos como Exercícios para o Pensamento*. Pelotas: IFSul, 2019. 142 p.

MONTOITO, R. *Uma Visita ao Universo Matemático de Lewis Carroll e o (Re)encontro com sua Lógica do Nonsense.* Dissertação (Mestrado em Educação) – Programa de Pós- Graduação em Educação. Universidade Federal do Rio Grande do Norte, Natal, 2007. 185 f.

MORTARI, C. A. *Introdução à lógica*. São Paulo: Editora UNESP, 2001. 393 p.

PITCHER, G. "Wittgenstein, Nonsense, and Lewis Carroll". *The Massachusetts Review,* v. 6, n. 3. 1965. p. 591-611.

PORFÍRIO. *Isagoge: Introdução às Categorias de Aristóteles*. Tradução: Bento Silva Santos. São Paulo: Attar, 2002. 95 p.

QUINE, W. O. *O Sentido da Nova Lógica.* 2 Ed. Curitiba: Ed. da UFPR, 1996. p. 15.

REICHERT, K. *Lewis Carroll: Studies zum literarischen Unsin*. Munique: Carl Hanser, 1974.

RUSSELL, B. *The Collected Papers of Bertrand Russell: Vol. 10*. Editado por John G. Slater. Londres: Routledge, 1996. p. 524.

SAUTTER, F. "As Teorias Carrollianas das Falácias". *Cadernos de História e Filosofia da Ciência,* Campinas, v. 1, n.1, jan./jun, 2015, p. 7-32.

SHAWYER, E. "Mrs. Shawyer's reminiscences". In : CARROLL, L. *Diaries*. GREEN, Roger Lancelyn (Editor). London: Cassell & Company LTD, 1953.

SILVEIRA, P. H. F. "A lógica do nonsense em Carroll e Wittgenstein". *Revista Filosofia, Ciência e Vida,* Ed. 49. Editora Escala, 2010.

SMITH, D; PROTEVI, J. "Gilles Deleuze". In: ZALTA, E. N. (Ed.). *The Stanford Encyclopedia of Philosophy*. Stanford, California. Spring, 2018. Disponível em: https://plato.stanford.edu/archives/spr2018/entries/deleuze/. Acesso em: 07 mar. 2020.

THE JOY OF LOGIC. Direção de Catherine Gale. 59 min. Londres: BBC Four, 2013. Documentário (59 min), son., col. Disponível em: http://www.-dailymotion.com /video/x1a6ogd_the-joy-of-logic_tech. Acesso em: 25 jan. 2020.

TOLKIEN, J. R. R. *Árvore e Folha*. 2 Ed. Tradução: Ronald Eduard Kyrmse. São Paulo: Editora WMF Martins Fontes, 2017. 106 p.

VALENCIA, V. S. "The Algebra of Logic". In: GABBAY, D. M.; WOODS, J. (Org.). *Handbook of The History of Logic: The Rise of Modern Logic from Leibniz to Frege.* Amsterdam: Elsevier B.V., 2004. p. 389-544.

WAGNER, D. "The uses of nonsense: Ludwig Wittgenstein reads Lewis Carroll". *Wittgenstein-Studien,* v. 3. 2012. p. 205-216.

WITTGENSTEIN, L. *Investigações Filosóficas*. Tradução: José Carlos Bruni. São Paulo: Abril S.A., 1975. 226 p. (Os pensadores)

WITTGENSTEIN, L. *Tractatus Logico-Philosophicus.* 3 Ed. Tradução: Luiz Henrique Lopes dos Santos. São Paulo: Editora da Universidade de São Paulo, 2001. 296 p.

WITTGENSTEIN, L. *Investigações Filosóficas/Philosophische Untersuchungen: Edição Bilíngue Alemão/Português*. Apresentação, tradução e notas de João José R. L. de Almeida. Campinas: Editora UNICAMP, [s./d.]. *No prelo*. 399 p. Disponível em: http://www.psicanaliseefilosofia.com.br/textos/InvestigacoesFilosoficas-Original.pdf. Acesso em: 28 mar. 2020.

AS LÓGICAS DOS TERMOS SINGULARES

Newton Marques Peron[1]

RESUMO

A partir de meados do século XIX, os termos singulares passaram a ter um papel crucial na filosofia da lógica e da linguagem. Este artigo se debruça sobre alguns problemas envolvendo termos singulares em cinco autores: Frege, Russell, Strawson, Kripke e Kaplan. Busco aqui, primeiramente, compreender de que modo a lógica desses filósofos está em consonância ou não com sua filosofia da linguagem no que diz respeito aos termos singulares. Como argumentarei no presente artigo, há três filósofos em que há uma dissintonia entre lógica e filosofia da linguagem: Frege, Strawson e Kaplan. No caso de Frege, analiso a proposta de Lehmann de uma lógica da filosofia da linguagem e concluo que ela é bastante adequada à visão fregeana. Nos casos de Strawson e Kaplan, defendo que as propostas existentes na literatura são inadequadas, e proponho uma alternativa de lógica da filosofia da linguagem para cada um desses dois filósofos. Por fim, esboço uma solução que consiste numa combinação de três tipos de lógica, de modo que tal combinação acomodaria grande parte das filosofias da linguagem quanto aos termos singulares dos cinco filósofos aqui tratados. Essa solução, além disso, dá conta de alguns problemas concernentes ao uso de nomes próprios em contextos ficcionais, algo que as abordagens dos filósofos acima listados deixam a desejar.

Palavras-chave: Termos Singulares. Nomes Próprios. Descrições Definidas. Nomes Ficcionais.

ABSTRACT

Since the middle of the 19th century, singular terms started to play a crucial role in the philosophy of logic and language. This paper focuses on some problems pertaining singular terms in five authors: Frege, Russell, Strawson, Kripke and Kaplan. Firstly, I try here to understand how the logic of those philosophers is or not in line with their philosophy of language concerning singular terms. As I will argue in this paper, there are three philosophers in which there is a disagreement between logic and philosophy of language: Frege, Strawson and Kaplan. In Frege's case, I analyze Lehmann's proposal for a logic of the philosophy of language and conclude that it is quite adequate to the Fregean view. In Strawson's and Kaplan's cases, I argue that the

[1] Professor permanente do Programa de Pós-Graduação em Filosofia da UFFS (Universidade Federal da Fronteira Sul), *campus* Chapecó.
E-mail: newton.peron@uffs.edu.br. ORCID: 0000-0003-3724-1775.

existing proposals in the literature are inadequate, and I propose an alternative logic of the philosophy of language for each of these two philosophers. Finally, I outline a solution that consists of a combination of three types of logic, so that such a combination would accommodate much of the philosophies of language regarding the unique terms of the five philosophers that are treated here. Moreover, this solution addresses some problems concerning the use of proper names in fictional contexts, something that the approaches of the philosophers listed above are found wanting.

Keywords: Singular Terms. Proper Names. Definite Descriptions. Fictional Names.

Introdução

Por mais de dois milênios, a lógica e a filosofia deram pouca importância aos denominados *termos singulares*. São exemplos de termos singulares: nomes próprios como "Sócrates", pronomes demonstrativos seguidos de termos gerais como "este filósofo" e descrições definidas como "o mestre de Platão".

Foi apenas em meados do século XIX, com os trabalhos de Frege, que os termos singulares passaram a ter um destaque especial na lógica e na filosofia da linguagem. Essa polêmica permeou grande parte do debate na filosofia analítica ao longo do século XX. Neste artigo, vamos nos concentrar em alguns desses debatedores, como Russell, Strawson, Kripke e Kaplan.

Com o intuito de apresentar de maneira esquemática a articulação entre lógica, filosofia da linguagem e termos singulares em cada um desses filósofos, consideraremos a classificação de Morris (2007, p. 69). Para o autor, com relação à articulação entre lógica e filosofia da linguagem na tradição analítica, há dois grandes grupos de filósofos.

O primeiro grupo vê o papel do filósofo como alguém que cria uma linguagem artificial para, a partir dela, erigir sua filosofia da linguagem. Esse grupo vê a linguagem humana como uma maquinaria complexa, cabendo ao filósofo desvendar o seu funcionamento. Nesse grupo, não há distinção entre filosofia da linguagem e lógica. Nesse artigo, pretendo apresentar boas razões para incluir nesse primeiro grupo Russell, Kripke e Kaplan.

Um segundo grupo, porém, é relutante em generalizar a riqueza da linguagem natural por meio de qualquer linguagem formal. Para esse grupo, a filosofia da linguagem não consiste em apresentar os mecanismos intrínsecos da linguagem natural por meio de um modelo formal, mas sim em compreender o que as pessoas fazem quando usam a linguagem. Aqui, o divórcio entre lógica e filosofia da linguagem é bastante natural e, acrescentaríamos, até esperado. Há, acredito, bons motivos para incluirmos Frege e Strawson nesse segundo grupo.

A divergência entre os dois grupos é muito menor, porém, do que a princípio possa parecer. Se entendermos algumas lógicas como um modelo formal para o papel dos termos singulares na filosofia da linguagem desses filósofos, teríamos – com todas as ponderações que devemos ter face a qualquer tabela esquemática em filosofia – a seguinte tabela:

Filosofia da Linguagem	**Lógica dos termos singulares**
Frege	Lógica Livre Neutra
Russell	Lógica Clássica
Strawson	Lógica Livre Neutra
Kripke	Lógica Modal Clássica
Kaplan	Lógica Híbrida Clássica.

Até agora, porém, não foi proposta nenhuma Lógica Híbrida Livre. Essa lógica, creio eu, com algumas adaptações pode ser entendida como um modelo formal capaz de acomodar grande parte das posições de Frege, Strawson, Kripke e Kaplan com relação aos termos singulares.

Acredito que podemos defender uma filosofia da linguagem que brote desse modelo formal e que não seja classificada em nenhum dos dois grupos descridos por Morris. Diferentemente de Russell, Kripke e Kaplan, que propõem uma interpretação definitiva dos termos singulares na linguagem natural que forçosamente se acomode a certo modelo formal ideal, defendo uma filo-

sofia da Lógica Híbrida Livre como uma maneira de produzir uma semântica formal que seja um mero modelo temporário para compreender um certo uso dos termos singulares na linguagem natural.

Além disso, nessa perspectiva não haveria um divórcio entre lógica e filosofia da linguagem, como há em Frege e Strawson. Mas isso porque a lógica não teria apenas o papel normativo de nos impor como devemos raciocinar, como é o caso da Lógica Clássica. No meu entender, a lógica tem muitos papéis e, dentre eles, o papel de fornecer um modelo – simplista e reducionista, como qualquer modelo – para compreender melhor um aspecto do que as pessoas fazem quando usam os termos singulares na linguagem natural.

1. Termos Singulares e a Revolução Fregeana

O conceito de termo singular nos remete à Grécia Antiga. Aristóteles, por exemplo, distingue em *Da Interpretação* (17a37-17b16) termos singulares de termos gerais. Os exemplos de Aristóteles para termos singulares são "Sócrates" e "este homem", que buscam se referir a um único indivíduo. Termos gerais são "homem" e "branco", que não se referem a um indivíduo particular, mas que buscam apreender certas características gerais de certos indivíduos.

Em *Primeiros Analíticos* (24a10-24a15), fica claro que a preocupação maior de Aristóteles não é o estudo das regras que formam as sentenças, mas certos tipos específicos de sentença: as proposições (*protasis*). Grosso modo, para Aristóteles proposições são sentenças que podem ser verdadeiras ou falsas. Mas dentre as proposições, apenas as proposições em que ocorrem somente termos gerais recebem sua atenção especial.

Tendo isso em vista, podemos nos arriscar a dizer que aquilo que posteriormente ficou conhecido como *lógica aristotélica* tinha como objeto de estudo sobretudo os termos gerais. Desse modo, durante séculos a distinção entre termos gerais e termos singulares tinha a função primordial de alertar que o objeto de investigação lógica no que concerne às partes da proposição deixaria em segundo plano os termos singulares.

Uma grande revolução ocorreu quando Frege, em *Fundamentos da Aritmética,* passou a defender que a análise lógica não deve se restringir aos

termos gerais, mas que devemos levar em conta também os termos singulares. Na verdade, essa mudança de enfoque é o reflexo linguístico de um abismo lógico aberto por Frege entre conceitos e objetos (1974, p. 202).

Enquanto o conceito é a referência (*Bedeutung*) de um termo geral (FREGE, 2009, p. 168), os termos singulares têm objetos como referência. Frege sustenta que, por um lado, há modos distintos de se referir a um mesmo objeto e, por outro, que a noção de objeto é muito ampla, de modo a abarcar não apenas indivíduos e objetos da nossa percepção sensível, mas também objetos não espaciais. Como exemplo de objetos não espaciais temos os números, que são objetos matemáticos (FREGE, 1974, p. 253).[2] É digno de menção que termos singulares abarcam não apenas nomes próprios, como "Aristóteles", mas também descrições definidas, como "o discípulo de Platão e mestre de Alexandre Magno". (FREGE, 2009, p. 132).

Como bem observou Frege, não há uma correspondência biunívoca nas linguagens naturais – como o alemão ou o português – entre termos singulares e objetos. Em outras palavras, não é verdade que nessas linguagens para cada objeto há um único termo singular que se refira a ele. De fato, podemos ter dois termos singulares se referindo a um único objeto. Esse é o caso dos termos singulares "a estrela da manhã" e "a estrela da tarde", que se referem ao mesmo objeto celeste, que é o planeta Vênus (FREGE, 2009, p. 131). Tampouco é verdade nessas linguagens que todo termo singular tem ao menos um objeto como sua referência. O termo singular "o corpo celeste mais distante da Terra", por exemplo, não se refere a objeto nenhum (FREGE, 2009, p. 133).

O fenômeno de termos singulares sem referência é mais evidente se olharmos para nomes próprios em contextos ficcionais, como "Ulisses". Visto que o personagem Ulisses da *Ilíada* e da *Odisseia* nunca existiu, "Ulisses" é um termo singular sem referência. Mas Frege explica que quando lemos lite-

[2] Estamos aqui excluindo e o Verdadeiro e o Falso, que são objetos lógicos (FREGE, 2009, p. 139). A razão disso é que o que estamos usando aqui a expressão "termos singulares" para englobar nomes próprios genuínos e descrições definidas, ou seja, partes de uma sentença, enquanto o Verdadeiro e o Falso são a referência de uma sentença assertiva completa (FREGE, p. 138). Outro modo de entender os termos singulares na linguagem natural em Frege é considerá-los como sendo tudo aquilo que é um nome de um objeto, excluindo-se o Verdadeiro e o Falso.

ratura, não estamos preocupados se Ulisses de fato existiu ou não e a questão da verdade ou falsidade das sentenças sequer se coloca. Quando, por outro lado, passamos a ter uma atitude de investigação científica, o valor de verdade da sentença passa a ter importância fundamental. Aí é importante garantirmos que cada termo singular tenha, ao menos, uma referência (FREGE, 2009, p. 137-138).

De acordo com Frege, se uma das partes de uma sentença assertiva completa não tem referência, a sentença como todo não pode ser considerada nem verdadeira nem falsa, ou seja, a sentença também não tem referência (2009, p. 138).[3] Esse princípio é conhecido na literatura como Princípio Compositivo (*Principle of Compositionality*), e pode ser anunciado do seguinte modo:

> **Princípio Compositivo (PC):** O valor semântico de uma expressão complexa depende do valor semântico de todas as expressões mais simples que ocorrem nela e do modo como essas expressões estão combinadas.[4]

[3] Para Frege, "sentenças assertivas completas" se referem ao Verdadeiro ou ao Falso e, portanto, essas sentenças poderiam ser entendidas como "nomes" desses objetos. Isso é o que aparece indicar em *Digressões Sobre o Sentido e Referência* (2009, p. 160): "Um nome próprio tem como referência o objeto que ele designa ou nomeia". Mas isso não valeria para qualquer sentença. Mesmo num contexto em que todos os termos singulares que ocorrem numa determinada sentença tenham referência, a sentença como um todo pode não se referir a um valor de verdade, por exemplo, se estamos diante de uma sentença subordinada. Nesse caso, a sentença pode se referir a um pensamento, a pedaços de pensamentos ou até mesmo a um indivíduo (2009, p. 142). No primeiro caso, Frege cita como exemplo a sentença "Copérnico acredita que as órbitas planetárias eram circulares". Aqui, a subordinada "que as órbitas planetárias eram circulares" não se refere a um valor de verdade (no caso, o falso, visto que as órbitas planetárias são elípticas), mas se refere a um pensamento de Copérnico: sua crença (equivocada) de que elas eram circulares (2009, p. 146). Já na sentença "Quem descobriu a forma elíptica das órbitas planetárias morreu na miséria", a subordinada "quem descobriu a forma elíptica das órbitas planetárias" não se refere a um pensamento nem a um valor de verdade, mas um indivíduo; nesse caso, a Kepler (2009, p. 146).

[4] Há diversas versões desse princípio na literatura. A versão aqui apresentada é uma adaptação da *Compositionality* da *Stanford Encyclopedia of Philosophy*. O autor da entrada, Szabó, usa o termo *meaning*, que deve ser melhor traduzido por "significado" (SZABÓ, 2020). Podemos entender, porém, o termo inglês *meaning* como englobando tanto o conceito fregeano de sentido (*Sinn*) como de referência (*Bedeutung*). De certo modo, há duas versões de **PC** endossadas por Frege: uma no que diz respeito ao sentido e outra no que diz respeito à referência. Formulamos aqui uma versão mais neutra do princípio por meio da expressão "valor semântico". Nossa intenção é compreender de que modo esse princípio pode ser ou não verificado nos demais filósofos tratados neste artigo, com visões semânticas distintas.

Isso significa que, no uso científico da linguagem, é importante que os termos singulares se refiram a objetos realmente existentes. Se isso acontecer, então podemos nos assegurar que as sentenças em que ocorrem esses termos singulares são verdadeiras ou falsas.[5] Não há nada na linguagem natural, porém, que garanta que os termos singulares se refiram a objetos realmente existentes.

A maneira de contornar essa debilidade das linguagens naturais no uso científico é criando uma nova linguagem, uma linguagem artificial. Eis a proposta fregeana: criar uma linguagem escrita de sinais, uma Conceitografia (*Begriffsschrift*). Nessa nova linguagem, todos os termos terão referência (FREGE, 2009, p. 132) e, portanto, toda sentença escrita nessa linguagem será ou verdadeira ou falsa.[6]

Em termos lógicos, essa linguagem deve respeitar ao menos dois preceitos:

P1 Toda sentença é verdadeira ou falsa, não podendo ser atribuídos a cada sentença ambos os valores de verdade ou nenhum;

P2 Todo termo singular se refere a um único objeto.

---

[5] Essa implicação é imprecisa. Além dos termos singulares terem referência, precisamos que os conceitos também tenham referência. Mas é importante ressaltar que algo diferente ocorre com os conceitos. Por exemplo, o termo singular "Pégasos" não tem referência porque Pégasos não existe. Mas o conceito "cavalo alado" tem referência e, como não existem cavalos alados, o conceito "cavalo alado" tem como extensão o conjunto vazio ou, em outras palavras, a função conceitual **cavalo_alado(x)** é falsa para qualquer valor de **x**. Para Frege, conceitos sem referência são conceitos imprecisos. Sobre conceitos vazios, ver, por exemplo, trecho de *Os Fundamentos da Aritmética* (1974, p. 262) em que se discute o conceito vazio "ser diferente de si mesmo"; sobre conceitos imprecisos, consultar *Digressões sobre Sentido e Referência* (2009, p. 166).

[6] A Conceitografia é uma linguagem escrita para expressar sentenças assertivas completas. Nessa linguagem, se duas sentenças contém o mesmo conteúdo conceitual, elas são indiscerníveis. Ela não distingue, por exemplo, as sentenças "Em Plateia, os gregos derrotaram os persas" e "Em Plateia, os gregos foram derrotados pelos persas". Além disso, na linguagem da Conceitografia todas as sentenças são consideradas nomes próprios. Isso é possível porque, por exemplo, em vez de se dizer "Aristóteles morreu em Siracusa" pode-se expressar o mesmo conteúdo conceitual dizendo-se "A morte de Aristóteles em Siracusa é um fato". Na linguagem criada por Frege, o único predicado comum a todas as sentenças é o símbolo $\vdash$, para marcar que a sentença formalizada que segue à direta do símbolo foi asserida (2012, p. 59-61). Ter uma linguagem constituída apenas de nomes próprios que se referem ao Verdadeiro ou ao Falso e sem distinção de conteúdo conceitual não são as únicas vantagens de uma Conceitografia. Em *Sobre a Justificação Científica de Uma Conceitografia*, Frege defende que, embora mais rica para outras finalidades, a linguagem natural falada contém uma variedade de modulações que são irrelevantes para o rigor das cadeias dedutivas. Outra vantagem da lin-

Vamos usar aqui a expressão "lógica do tipo clássico" num sentido bem amplo e grosseiro, de modo que definiremos lógica do tipo clássico como sendo toda linguagem artificial que respeite **P1**, **P2** e **PC**. Nesse sentido lato, a Conceitografia é claramente uma lógica do tipo clássico.

Frege, porém, defende que **P1** e **P2** não são respeitados nas linguagens naturais. Na verdade, se admitirmos o Verdadeiro e o Falso com sendo referências das sentenças e, ao mesmo tempo, aceitarmos **PC**, temos que toda linguagem que não respeitar **P2** também não respeitará **P1**. E, pela ocorrência de termos ficcionais, é evidente que as linguagens naturais não respeitam **P2** e, portanto, também não respeitam **P1**.

1.1. A Lógica da Filosofia da Linguagem de Frege

Considerando que na visão de Frege as linguagens naturais não respeitam **P1** e **P2**, é fácil concluirmos que a lógica das linguagens naturais de Frege não é uma lógica do tipo clássico em sentido amplo. Qual seria, pois, a lógica das linguagens naturais do ponto de vista da filosofia da linguagem de Frege? Segundo Lehmann, em seu artigo *Strict Fregean Free Logic*, a lógica das linguagens naturais desse ponto de vista – o que o autor denomina de Lógica Fregeana em Sentido Estrito – é um tipo de Lógica Livre.

Lógicas livres são, grosseiramente falando, linguagens artificiais que não respeitam **P2** na medida em que nessas lógicas há termos sem referência, os denominados termos vazios. *Lógicas livres positivas*, por exemplo, são lógicas em que se um termo singular **t** não tiver referência, sentenças do tipo "**t** é **P**" podem ser verdadeiras. Já *lógicas livres negativas* são aquelas em que a ocorrência de um termo singular **t** sem referência força que sentenças do tipo "**t** é **P**" sejam falsas (NOLT, 2020). Daqui, podemos inferir que tanto a lógica livre positiva quanto a negativa respeitam **P1** mas não respeitam **P2**, portanto, tampouco respeitam **PC**.

---

guagem escrita à falada é sua invariância e permanência. Essas vantagens são mais bem exploradas, por exemplo, pela linguagem da aritmética. Para Frege, esse tipo de linguagem expressa as coisas diretamente, sem precisar da medição sonora. Mas a linguagem aritmética não consegue expressar conexões lógicas. Daí a necessidade de uma Conceitografia, na acepção estrita to termo (2009, 63-65).

A lógica de Frege, segundo Lehmann, não é uma lógica livre positiva nem negativa. É o que o autor denomina lógica NINO, abreviação inglesa para *No Input, No Output* (LEHMANN, 1994, p. 310) e que, mais tarde, passou a ser considerada um tipo de *Lógica Livre Neutra* (NOLT, 2020). Em NINO, **PC** é respeitado, de modo que se um termo sem referência ocorre numa sentença, a sentença como um todo não terá valor de verdade; porém, se todos os termos que ocorrem numa sentença tiverem referência, a sentença como um todo será ou verdadeira ou falsa. Como toda lógica livre, a lógica de Lehmann não respeita **P2**; e, por ser uma Lógica Livre Neutra, a lógica de Lehmann também não respeita **P1**.

Vimos que, para Frege, as linguagens naturais respeitam **PC** e, ao mesmo tempo, não respeitam **P1** nem **P2**. Daqui, podemos inferir que a Lógica Livre Neutra de Lehmann é um modelo bastante fidedigno à filosofia da linguagem de Frege. Ou, se formos mais cautelosos, podemos assegurar que a lógica de Lehmann está muito mais próxima da filosofia da linguagem fregeana do que as demais alternativas vistas até então: lógica do tipo clássico, a lógica livre positiva e a lógica livre negativa.

Concluímos, portanto, que há ao menos dois sentidos de "lógica fregeana". No primeiro sentido, é a linguagem artificial proposta por Frege. Essa lógica fregeana é claramente a Conceitografia, que respeita **P1**, **P2** e **PC**. Num segundo sentido, podemos entender como "lógica fregeana" a linguagem artificial que pretende ser um modelo formal de sua concepção filosófica da linguagem. E é justamente nesse segundo sentido que a lógica fregeana é a Lógica Livre Neutra de Lehmann.

2. Russell e as Descrições Definidas

O problema da abordagem fregeana dos termos singulares é que ela nos leva ao seguinte paradoxo. Visto que Ulisses é um nome ficcional, então a sentença

(1) Ulisses não existe

deveria ser verdadeira. Mas como, para Frege, "Ulisses" é um termo singular sem referência, temos, por **PC,** que a sentença (1) não pode ser nem verdadeira nem falsa. Aqui, estamos diante do seguinte paradoxo fregeano com relação aos nomes próprios ficcionais: se Fulano de Tal nunca existiu, então a sentença "Fulano de Tal não existe" não é verdadeira (mas também não é falsa).

A solução de Russell em *Da Denotação* é entender as condições de verdade da sentença (1) como condições verdade de um outro tipo de sentença, em que uma certa descrição definida substitui o nome próprio Ulisses em (1). Descrições definidas são expressões que começam com um artigo definido singular da forma: "o(a) **x** tal que …" (1974b, p. 11).

Os artigos definidos singulares "o" e "a" são, na visão russelliana, abreviações de uma sentença complexa envolvendo quantificadores.[7] A função lógica desses artigos definidos singulares é garantir a existência e unicidade. Assim, para Russell, a sentença (1) deve ser entendida como:

> (1r) Não é o caso que existiu um único[8] **x** tal que **x** era rei de Ítaca, **x** lutou na Guerra de Troia…

Aqui, as reticências devem ser preenchidas por todas as demais características de Ulisses descritas por Homero na *Ilíada* e na *Odisseia*.[9] Como *Ilía-*

[7] Isso quer dizer que contextos sentenciais em que ocorrem descrições definidas devem ser parafraseados em contextos em que essas descrições definidas não ocorrem. Para isso, usamos quantificadores, operadores lógicos e a igualdade.

[8] Na linguagem formal de Russell, o quantificador "existe um único **x**" não é primitivo. Como símbolos primitivos temos: (i) o quantificador universal "para todo **x**"; (ii) o operador unário "não"; (iii) o operador binário "se… então"; e (iv) a relação binária "**x** é igual a **y**". Com o quantificador universal e a negação, podemos claramente definir o existencial. Por exemplo, a sentença "existe ao menos um **x** que tem a propriedade **P**" é equivalente, por definição, a "não é o caso que para todo **x**, **x** não tem a propriedade **P**". Em seguida, podemos expressar a sentença "existe um único **x** que tem a propriedade **P**" do seguinte modo: "existe ao menos um **x** que tem a propriedade **P** e, para todo **y**, se **y** tem a propriedade **P**, então **y** é igual a **x**". A primeira parte da sentença garante que existe ao menos um **x**, enquanto a segunda parte garante que não existe mais de um **x** (1974b, p. 10-11).

[9] Em *A Filosofia do Atomismo Lógico*, Russell usa como exemplo a sentença "Rômulo existiu", em que o nome Rômulo deve ser substituído por "existiu um único **x** tal que **x** tem tais e tais propriedades", tendo em vista todas aquelas propriedades que encontramos nos textos do historiador romano Títio Lívio (1974a, p. 111). Em *Introdução à Filosofia Matemática*, o mesmo método é usado para decidir o valor de verdade da sentença "Homero existiu": deve-se determinar se é verdadeira a sentença "existiu um único **x** tal que **x** escreveu a *Ilíada* e a *Odisseia*" (1996, p. 213).

da e a *Odisseia* são obras ficcionais, sabemos que (1r) é verdadeira. Mas (1r) expressa as condições de verdade de (1). Desse modo, como (1r) é verdadeira, (1) também é verdadeira, como esperávamos. E isso, segundo Russell, eliminaria o paradoxo da abordagem de Frege dos termos singulares.

Para compreendermos melhor a teoria de Russell, vamos aplicá-la ao seguinte exemplo de Frege:

(2) Ulisses desembarcou em Ítaca.

Como vimos, para Frege a sentença (2) não é nem verdadeira nem falsa, pois "Ulisses" não tem referência. Para Russell, porém, a sentença (2) deve ser analisada do seguinte modo:

(2r) Existiu um único **x** tal que **x** era rei de Ítaca, **x** lutou na Guerra de Troia e … e esse **x** desembarcou em Ítaca.

A sentença (2), para Russell, não deve ser considerada sem valor de verdade, como acreditava Frege, mas falsa. A razão para isso é a seguinte. Sabemos que a *Ilíada* e a *Odisseia* tratam-se de obras ficcionais e, portanto, nunca existiu um indivíduo que realizou todos os feitos que Homero atribui a Ulisses. Desse modo, a expressão "**x** era rei de Ítaca, **x** lutou na Guerra de Troia..." é falsa para qualquer valor de **x**. Daí, concluímos que a sentença (2r) é falsa. Mas (2r) expressa as condições de verdade de (2). Portanto, como (2r) é falsa, (2) também é falsa.

Tomemos agora a seguinte sentença:

(3) Ulisses não desembarcou em Ítaca

Para Frege, (3) também não é nem verdadeira nem falsa, desrespeitando **P1**. Mas para Russell (1974b, p. 17-18), tal sentença é ambígua. Desse modo, há duas maneiras de analisar as condições de verdade de (3), a saber:

(3r) Existiu um único **x** tal que **x** era rei de Ítaca, **x** lutou na Guerra de Troia e … e esse **x** não desembarcou em Ítaca;

(3r*) Não existiu um único **x** tal que **x** era rei de Ítaca, **x** lutou na Guerra de Troia e … e esse **x** desembarcou em Ítaca.

Visto que Ulisses nunca existiu, então (3r) é falsa e (3r*) é verdadeira.

Uma das vantagens da abordagem de Russell é que ela respeita **P1** mesmo para aquelas sentenças em que ocorrem nomes ficcionais. De fato, enquanto Frege entende que nomes ficcionais são termos singulares sem referência, Russell defende que sentenças em que ocorrem tais termos devem ser reescritas de modo a eliminar tais termos, obtendo, assim, sentenças logicamente equivalentes envolvendo existência e unicidade. Logo, toda sentença em que ocorrem termos singulares será ou verdadeira ou falsa, independentemente desses termos singulares terem ou não referência, que é o que exige **P1**.

Além disso, **PC** é respeitado na perspectiva russeliana pela seguinte razão. Sentenças em que ocorrem termos singulares são, em geral, reescritas de modo a obter sentenças logicamente equivalentes em que ocorre o operador existencial e a igualdade. A sentença reescrita, por sua vez, será verdadeira caso o termo singular tenha referência ou falsa caso o termo singular não tenha referência. Além disso, calcular o valor de verdade de uma sentença em que ocorre um termo singular é o mesmo que calcular o valor de verdade de uma sentença complexa cujo pedaço é uma descrição definida, ou seja, cujo pedaço também é uma sentença, mas de complexidade menor. Ora, se quisermos calcular o valor de verdade da sentença mais complexa, temos que calcular o valor de verdade de um pedaço menos complexo primeiro para, em seguida, calcular o valor de verdade da sentença como um todo, que é mais complexa. Assim, a verdade ou a falsidade dependem tanto da referência dos termos singulares que nela ocorrem quanto da verdade ou falsidade das sentenças componentes e do modo como estão combinadas, que é o que afirma **PC**.[10]

[10] A verdade e falsidade devem ser pensadas no vocabulário russeliano não como objetos lógicos, mas como um modo pelo qual certa proposição aponta para um fato correspondente

A abordagem de Russell também respeita **P2**, mas de uma maneira indireta. Isso porque Russell entende que grande parte daquilo que foi denominado pela tradição de termos singulares é, na verdade, sentenças envolvendo existência e unicidade. Além de descrições definidas e termos ficcionais, outros termos devem ser entendido como sentenças. Esse é o caso, por exemplo, de nomes próprios como "Apolo", que pode ser entendido como uma abreviação da descrição definida "o deus Sol" (RUSSELL, 1974b, p. 18-19). Mesmo nomes próprios de seres humanos reais são, em sua maior parte, compreendidos como descrições definidas (RUSSELL, 1996, p. 213). O nome próprio "Aristóteles", por exemplo, poderia ser entendido por alguém como uma abreviação da descrição definida "o discípulo de Platão e preceptor de Alexandre Magno e…" em que as reticências abreviam certas propriedades que nós atribuímos a Aristóteles.[11]

Então quais expressões da linguagem natural seriam candidatas a legítimos termos singulares para Russell? Isso depende do modo como conhecemos o objeto (RUSSELL, 1974b, p. 20). Se nós tivermos um *conhecimento de trato* (*acquaintance*) do objeto, então usamos um termo singular genuíno para nos referimos a este objeto. Grosso modo, temos um conhecimento de trato de um objeto quando tal conhecimento não ocorre por meio de descrições definidas, mas por meio de um acesso direto ao objeto. Isso acontece, basicamente, em duas situações. No primeiro caso, quando usamos nosso nome próprio para nos referirmos a nós mesmo, como a sentença "Aristóteles é filósofo" dita pelo próprio Aristóteles (RUSSEL, 2001 p. 62). No segundo caso, quando nos referirmos a algo que está no nosso campo de percepção. No caso do português, a maneira mais comum de nos referirmos a algo que está próximo a nós é por meio dos pronomes "este", "esta" e "isto", como em "isto é branco" (RUSSELL, 1974a, p. 77).

Quando falamos de nós mesmo, temos a garantia que nosso nome se refere a um único indivíduo e que, portanto, tem uma única referência. É o

(RUSSELL, 1974a, p. 84). De qualquer modo, a expressão "valor semântico" abarca o verdadeiro e o falso não só como objetos lógicos, mas também como correspondência entre proposições e fatos. Agradecemos ao parecerista deste artigo pela perspicaz observação.

[11] De acordo com Russell, a propriedades atribuídas na descrição definida que substitui o nome próprio "Aristóteles" dependem do grau de intimidade que a pessoa que usa o nome "Aristóteles" tem com Aristóteles (Russell, 2001, p. 63)

conhecimento de trato de nossa autoconsciência que garante referência ao nosso nome próprio nessa situação (RUSSELL, 2001, p. 58). Além disso, quando estamos diante de algo do nosso campo de percepção, estamos diante de um particular. Nesse caso, ao usarmos os pronomes "este", "esta" e "isto", temos também a garantia de estamos nos referindo a um único objeto existente (RUSSELL, 1974a, p. 78). Podemos concluir daqui que o conhecimento de trato dos termos singulares genuínos garantem que a filosofia da linguagem de Russell também respeite **P2**.

2.1. Há Uma Lógica da Filosofia da Linguagem em Russell?

Enquanto em Frege há uma distinção nítida entre a lógica apresentada em sua Conceitografia e a lógica de sua concepção filosófica da linguagem, a linha divisória se torna muito mais tênue no caso de Russell. Há, acredito, duas razões para isso.

A primeira razão de Russell não separar lógica de filosofia da linguagem natural é a seguinte. Tanto Frege quanto Russell viam numa lógica do tipo clássico a melhor maneira de reformular a matemática de modo a abandonar as imprecisões e ambiguidade que ocorrem quando os enunciados matemáticos são apresentados em linguagem natural. Mas sabemos que a linguagem matemática permite falar de termos singulares que não se referem a nada, por exemplo, o termo singular "a divisão de 1 por 0" não se refere a número algum. De acordo com o próprio Russell, a solução dada por ele anteriormente a esse problema havia sido a mesma defendida por Frege*:* atribuir uma referência arbitrária a todos esses termos matemáticos que não se referem a nada.

Mas no momento em que escreve *Da Denotação,* Russell admite que a solução de Frege não o satisfaz e apresenta essa como sendo uma das razões para adotar sua teoria das descrições definidas para lidar com o problema de termos singulares sem referência na linguagem natural (1974b, p. 10). Ou seja, sua nova solução para os termos singulares vazios na linguagem formal da lógica matemática é automaticamente aplicada à linguagem natural, sem qualquer mediação. Esse é um bom motivo para admitirmos que, para Russell,

não havia distinção alguma entre a linguagem formal adequada para o contexto matemático e a linguagem formal como mecanismo de análise de expressões linguísticas no seu uso cotidiano.

Além disso, cabe lembrar que um dos motivos de Russell preferir sua abordagem em detrimento à abordagem de Frege dos termos singulares na linguagem natural é porque ela preserva o Princípio do Terceiro Excluído (*law of excluded middle*), que é o princípio pressuposto em **P1**.[12] Mas esse princípio não é necessariamente um princípio da linguagem natural em seu uso cotidiano, e sim um princípio de uma lógica de tipo clássico. Russell está aqui, mais uma vez, defendendo que uma lógica do tipo clássico é o mecanismo mais adequado para análise da linguagem natural nos seus mais diversos usos.

Entendo, desse modo, que há sim uma lógica da filosofia da linguagem de Russell, mas que essa lógica é a própria versão russeliana da Lógica Clássica.

3. Strawson Contra Russell

A teoria das descrições definidas de Russell parecia ter dado uma resposta definitiva ao problema dos termos singulares sem referência e, durante cerca de cinquenta anos, se tornou praticamente consensual entre os filósofos analíticos. A polêmica foi reacendida por Strawson em seu artigo *Sobre o Referir*.

Para Strawson, o problema da teoria das descrições de Russell é que ela não está de acordo com o uso corrente da linguagem. Para defender esse ponto de vista, Strawson considera o seguinte exemplo:

(5) O atual rei da França é sábio.

[12] Em (RUSSELL, 1974a) traduz-se a expressão original *law of excluded middle* (RUSSELL, 1905, p. 485) por "lei da exclusão dos meios". Embora literal, essa tradução não está de acordo com a expressão normalmente usada em manuais de lógica em língua portuguesa. Vide (MORTARI, 2016, p. 247) e (HEGENBERG, p. 57).

Na visão de Russell, a descrição definida "o atual rei da França" deve ser entendida como "existe um único **x** tal que **x** é rei da França". Desse modo, a sentença (5) deve ser equivalente em significado a:

(5r) Existe um único **x** tal que **x** é rei da França e **x** é sábio.

Tendo em vista que uma parte da sentença expressa em (5r), a saber, "**x** é rei da França" é falsa para todo **x**, temos que a sentença (5r) como um todo também é falsa. Mas se perguntarmos a alguém na rua se é verdade que o rei da França é sábio, a pessoa provavelmente ficará muito encabulada e não saberá responder. Como enfatiza Strawson, em vez de dizer que (5) é simplesmente falsa, a pessoa pedirá para você repetir a pergunta ou dirá algo do tipo "você cometeu um equívoco, não há mais rei na França". Por outro lado, se (5) fosse proferida (*uttered*) em vários momentos durante o século XVIII, ela poderia ser verdadeira ou falsa, dependendo a que monarca o falante estivesse se referindo (STRAWSON, 1974, p. 270-271).

Daqui, Strawson defende que devemos distinguir entre a utilização de uma expressão e o proferimento de uma expressão. Se duas pessoas usarem a expressão "o atual rei da França" em tempos diferentes, elas podem estar se referindo[13] (*referring*) a monarcas distintos. Nesse caso, são duas utilizações distintas da mesma expressão. Se essas duas pessoas estiverem pronunciando a expressão ao mesmo tempo, estamos diante da mesma utilização dessa expressão, mas em dois proferimentos distintos.

Porém, quando (5) é proferida em 1950, a expressão "o atual rei da França" não se refere a nada e (5), portanto, não é nem verdadeira nem falsa, pois a questão da verdade ou falsidade de (5) sequer pode ser colocada. Strawson ainda defende na mesma obra que nomes próprios ficcionais também não

[13] Russell em *Da Denotação* espelha a distinção fregeana entre *Sinn* e *Bedeutung* como *meaning* e *denotation* (1905, p. 483). Já Strawson usa *significance* como sinônimo de *meaning* e *referring* como sinônimo de *denotation* (1950, p. 323). Para aproximar o diálogo entre esses autores, estamos aqui traduzindo *meaning* e *significance* por "significado" enquanto os três termos *Bedeutung*, *denotation* e *referring* traduzimos indiscriminadamente por "denotação" ou "referência". Sabemos, porém, que os conceitos de *denotation* e *referring* não são coextensivos em inglês. Conferir o artigo de Donellan intitulado *Reference and Definite Description* (1966, p. 281-282).

têm referência. Desse modo, ele concorda com Frege que a sentença (2) não tem valor de verdade, visto que "Ulisses" não tem referência.

À primeira vista, a solução de Strawson não parece resolver sentenças do tipo (1). Mas no final de seu artigo, Strawson insiste que em proposições do tipo "Algum homem é mortal" está pressuposta a existência de seres humanos para ela ser verdadeira ou falsa. Se aplicarmos o mesmo raciocínio de Strawson não apenas para termos gerais mas também para termos singulares, diríamos que a existência de Ulisses está pressuposta para (1) ser verdadeira ou falsa e que, portanto, a questão da verdade ou falsidade de (1) sequer deve ser levantada em contextos ficcionais.

Outra maneira de contornar a dificuldade é dizer que (1) não expressa uma proposição justamente porque é usada em um contexto ficcional. Como defende Caorsi em seu texto *Referência e Termos Singulares*, Strawson parece oscilar entre essas duas posições não apenas nesse artigo mas também em outras obras posteriores (CAORSI, 2005, p. 179-181). Strawson replica Caorsi, enfatizando que, na verdade, essas duas posições não são mutualmente incompatíveis e mostra como reconciliá-las (STRAWSON, 2014, p. 83).

3.1. As Lógicas da Filosofia da Linguagem de Strawson

É verdade que Strawson nunca apresentou uma lógica formal para sua filosofia da linguagem. De fato, Strawson admite, em sua réplica a Caorsi, que o problema da lógica subjacente à sua filosofia da linguagem nunca o atraiu (2014, p. 84). Além disso, o seu texto *Sobre o Referir* finaliza concluindo que "nem as regras aristotélicas nem as russellianas fornecem a lógica exata de qualquer expressão da linguagem ordinária, pois a linguagem ordinária não tem uma lógica exata" (1974, p. 286).

Em Russell, o problema da lógica exata subjacente a sua filosofia da linguagem sequer é colocado, em virtude de haver uma identidade entre sua lógica e sua filosofia da linguagem. Já tal problema pode sim ser colocado concernente à filosofia da linguagem de Strawson por, ao menos, dois motivos. Primeiramente porque, como vimos, o próprio Strawson admite que há uma separação entre lógica formal e lógica da linguagem natural. O segundo moti-

vo é que existe, na literatura, ao menos duas tentativas de abarcar a filosofia da linguagem de Strawson em termos formais. E essas duas tentativas usam lógicas que não respeitam **P1** e **P2**, verificados em qualquer lógica do tipo clássico.

A primeira proposta, esboçada por Caorsi, é pensar uma lógica proposicional em que há um operador binário de pressuposição distinto da implicação material.[14] Caorsi propõe a operação "**p** pressupõe **q**" de modo que "**p** pressupõe **q**" é verdadeiro quando a verdade de **q** é uma condição para a verdade ou falsidade de **p** (CAORSI, 2015, p. 180-181).

Caorsi, porém, não diz se esse operador binário de pressuposição deve ser definido numa lógica do tipo clássico ou não. Desse modo, cabe perguntar: Strawson está defendendo que é possível definir, além da implicação material, uma implicação de pressuposição numa lógica proposicional de tipo clássico? Esse não parece ser o caso. A alternativa – que parece ter sido indicada pelo próprio Strawson em sua réplica a Caorsi – é tomar uma lógica proposicional de três valores, em que o terceiro valor representaria o indeterminado (2014, p. 84). Uma lógica assim claramente não respeitaria **P1**.

Existem, na literatura, várias lógicas proposicionais de três, quatro e até infinitos valores de verdade. Em cada uma dessas lógicas, podemos definir mais de uma implicação. Mas não vejo como definir a operação "**p** pressupõe **q**" em qualquer uma dessas lógicas e a razão é a seguinte. Segundo a perspectiva de Caorsi endossada por Strawson, se (5) receber o valor indeterminado é porque (5r) é falso. Isso significa que o valor de verdade de (5r) depende do valor de verdade de (5) para que "(5) pressupõe (5r)" seja verdadeiro. Daqui, se segue que a relação "**p** pressupõe **q**" não é uma função proposicional.[15] Desse modo, a relação "**p** pressupõe **q**" não pode ser definida em qualquer ló-

[14] A implicação material é uma implicação definível em qualquer lógica proposicional do tipo clássico. As condições de verdade de uma implicação material são: dadas duas proposições **p** e **q**, temos que "**p** implica **q**" é falso apenas no caso em que **p** é verdadeiro e **q** é falso; nos demais casos, "**p** implica **q**" é verdadeiro.

[15] Talvez esse ponto fique mais claro com uma analogia. A relação "a soma de **x** com **y**" é uma operação numérica da aritmética porque a função tem apenas os valores de **x** e de **y** como entrada, de modo que tanto **x** como **y** se referem a números. Já a relação "**p** pressupõe **q**" não pode ser uma operação proposicional. De fato, quando "**p** pressupõe **q**" é verdadeiro, tanto **p** quanto **q** são sentenças envolvendo quantificadores.

gica proposicional, independentemente de quantos valores de verdade essa lógica viesse a ter.

Buscando mudar a lógica no que diz respeito não à implicação mas à referência dos termos singulares, van Frassen considerou uma lógica livre neutra diferente da de Lehmann como a lógica da filosofia da linguagem de Strawson com relação ao tratamento dos termos singulares (1966, p. 181). A ideia fundamental de van Fraasen é que sentenças em que ocorrem expressões sem referência não têm valor de verdade, mas seria possível atribuir certas referências a essas expressões expandido o domínio de objetos.

A sugestão de van Fraasen foi desenvolvida tecnicamente alguns anos mais tarde por Bencivenga. Vamos denominar de *domínio original de objetos* aquele domínio em que, numa certa linguagem, há termos sem referência. Há muitas maneiras de expandir esse domínio original, atribuindo referências distintas aos termos que, outrora, não tinham referência. Uma sentença é verdadeira quando para toda expansão do domínio original a sentença é verdadeira, e uma sentença é falsa quando para toda expansão do domínio original a sentença é falsa. Quando a sentença é verdadeira para algumas expansões do domínio original e falsa para outras expansões, a sentença considerando o domínio original não é nem verdadeira nem falsa (BENCIVENGA, 1986, 9. 409). A lógica de Bencivenga, sendo uma lógica livre, claramente não respeita **P2**. E sendo uma lógica livre neutra, também não respeita **P1**.

A meu ver, essa solução se desvia um pouco da visão strawsoniana. De acordo com Strawson, a sentença (5) pode ser: (i) verdadeira se dita durante o reinado de Luís XIV; (ii) falsa se dita durante o reinado de Luís XV; (iii) sem valor de verdade se dita em 1950. Estamos aqui diante de três épocas distintas. Digamos que cada expansão do domínio original seja a contraparte formal daquilo que Strawson denomina *época* (*time*). Assim, a referência do termo singular "o atual rei da França" mudaria de época para época. A contrapartida desse fenômeno na linguagem formal de Bencivenga seria a seguinte: a referência de um termo singular **t** mudaria em cada expansão do domínio original (se expansões distintas do domínio original não forem a contraparte formal na lógica de Bencivenga para épocas distintas, não sei o que seria).

Mas isso implica que na lógica de Bencivenga a sentença (5) seria sempre ou verdadeira ou falsa em cada expansão específica do domínio. E se a expansão do domínio é a contraparte formal de uma determinada época, teríamos que (5) seria sempre verdadeira ou falsa em cada época específica, apesar de não conseguirmos determinar a verdade ou a falsidade de (5) independentemente da época. Mas não é isso que diz Strawson: há épocas em que (5) não é nem verdadeira nem falsa, a saber, na época em que o termo "o atual rei da França" não se refere a ninguém. Os modelos de Bencivenga, porém, diriam outra coisa: a sentença (5) não é "eternamente" nem verdadeira nem falsa porque em certas épocas (5) é verdadeira e em outras épocas (5) é falsa, mas não haveria uma época em que (5) não fosse nem verdadeira nem falsa.

Isso nos sugere considerar a lógica NINO de Lehmann mais próxima da filosofia da linguagem de Strawson do que a de Bencivenga se não fosse pela seguinte limitação. Em NINO, não há nada que possa interpretar o termo "atual" na própria linguagem formal e, portanto, fica difícil considerar NINO um modelo apropriado daquilo que seria a época de enunciação de (5) para Strawson.

Aqui, a palavra-chave parece ser "atual", que pertence a uma certa categoria de palavras que Kaplan denominou *indexicais*. Além disso, como veremos adiante, a filosofia da linguagem de Kaplan parece ter uma contraparte formal muito adequada nas denominadas lógicas híbridas. Mas para compreendermos, em linhas gerais, a filosofia da linguagem de Kaplan e os modelos de lógica híbrida, é preciso retrocedermos um pouco e compreendermos alguns conceitos básicos da filosofia da linguagem de Kripke e a sua lógica modal.

4. Kripke e os Designadores

Para Frege, nomes próprios são termos singulares que têm mais de um sentido (*Sinn*). Esse sentido normalmente é expresso por meio de uma descrição definida. Por exemplo, um dos sentidos do nome próprio "Vênus" é "a estrela da manhã", outro sentido do mesmo nome próprio é "a estrela da tarde".

Ambos os sentidos se referem ao mesmo objeto, que é um certo corpo celeste no nosso sistema solar.

Russell, por sua vez, discorda que nomes próprios sejam termos singulares cujo sentido é uma certa descrição definida, podendo ou não se referir a um certo objeto. Para ele, nomes próprios têm, em geral, a mesma função que descrições definidas. Do ponto de vista lógico, sentenças envolvendo descrições definidas devem ser reescritas de modo a obter uma sentença equivalente envolvendo existência e unicidade. Se há uma única denotação para o nome próprio, a sentença é verdadeira; caso contrário, ela é falsa.

Mas há algo comum nessas duas perspectivas. Em ambas, há uma relação muito estrita entre nomes próprios da linguagem natural e descrições definidas. Em Frege, a descrição definida determina o sentido de um nome próprio; em Russell, a descrição definida determina o valor de verdade de uma sentença em que ocorre um nome próprio.

Kripke, em *O Nomear e a Necessidade,* se contraporá àquilo que ele denomina de perspectiva de Frege e Russell sobre os termos singulares, ou seja, "designadores" no vocabulário kripkeano (2012, p. 70).[16] Grosseiramente, essa perspectiva pode ser resumida segundo Kripke do seguinte modo: para Frege, descrições definidas são cruciais para determinar o sentido/significado de um nome; para Russell, elas são fundamentais para determinar a denotação/referência de um nome. Kripke defende que descrições definidas não têm nenhum papel para determinar quer o significado quer o referente do nome. Além disso, Kripke dirá que um nome próprio sequer têm sentido/significado, mas se refere sempre diretamente ao objeto portador do nome.

Para defender sua tese, Kripke usa argumentos modais, ou seja, análises de sentenças da linguagem natural em que ocorrem termos modais como "necessário", "contingente" e "possível". Vejamos um exemplo do próprio Kripke (2012, p. 110):

[16] Para sermos fiéis ao trecho citado, o conceito de "designador" engloba apenas nomes próprios e descrições definidas, mas não engloba expressões como "este homem", em que há um pronome seguido de um termo geral. Porém, cabe mencionar que, em outra passagem, Kripke sugere que pronomes também devam ser compreendidos como designadores (2012, p. 51).

(6) Aristóteles poderia não ter sido discípulo de Platão.

Sabemos que (6) é verdadeira. Mas se o nome próprio "Aristóteles" for sinônimo da descrição definida "o discípulo de Platão e mestre de Alexandre Magno", então (6) seria contraditória, pois seria equivalente à sentença "o discípulo e Platão e mestre de Alexandre poderia não ter sido discípulo de Platão".

Pela nossa intuição, (6) não é contraditória; e sim, verdadeira. Desse modo, o nome próprio "Aristóteles" não deve ser entendido com sinônimo da descrição definida "o discípulo de Platão e mestre de Alexandre Magno". O argumento modal pode ser facilmente generalizado para qualquer descrição definida atribuída ao nome "Aristóteles", por maior que ela fosse. Fica clara, assim, a limitação da teoria das descrições definidas dos nomes próprios como teoria que busca determinar o sentido/significado dos nomes próprios.

Esse argumento pode ser facilmente adaptado para o problema de determinar o valor de verdade da sentença (6). Pela nossa intuição, (6) é verdadeira. Vimos, porém, que pela teoria das descrições definidas, a sentença (6) é contraditória, ou seja, verdadeira e falsa ao mesmo tempo. Considerando que o valor de verdade é a referência de uma dada sentença, temos que o argumento modal mostra também um problema na teoria das descrições definidas dos nomes próprios como teoria que determina a referência/denotação dos nomes próprios.

Kripke enfatiza que, para determinar o valor de verdade de (6), temos que pensar numa situação contrafactual. Nessa situação contrafactual, que ele denomina de um outro mundo possível, Aristóteles teria tido qualquer outra ocupação diferente de fazer filosofia. Nesse mundo possível, usamos o nome próprio "Aristóteles" para nos referirmos ao mesmo indivíduo Aristóteles. Já a descrição definida "o discípulo de Platão e mestre de Alexandre Magno" poderia se referir nesse mundo, por exemplo, a Espeusipo (que também foi discípulo de Platão no mundo atual mas que não foi mestre de Alexandre o Grande no mundo atual) ou simplesmente a ninguém.

Para se contrapor à perspectiva descritivista, Kripke defende duas principais teses. A primeira tese é a de que nomes próprios são designadores rígidos porque se referem ao mesmo indivíduo ao longo dos mundos possíveis. Já descrições definidas são designadores não-rígidos porque podem se referir a indivíduos distintos ao longo dos mundos possíveis (2012, p. 98-99).

A segunda tese é a de que, na maioria dos casos, a referência dos nomes próprios não é fixada por meio de descrições definidas, mas sim a partir de um batismo inicial.[17] Em seguida, o nome espalha-se de elo em elo numa cadeia comunicativa dentro de comunidade de falantes.

4.1. As Lógicas Modais de Kripke

Nota-se que todas as objeções de Kripke à teoria descritivista estão alicerçadas sobre uma maneira de interpretar as expressões modais "necessário" e "possível". Em termos mais técnicos, todas as objeções pressupõem uma semântica formal que interprete os operadores modais "necessário" e "possível".

Como devemos fazer para determinar o valor de verdade de uma sentença com um operador modal, como a sentença (6)? A solução que Kripke encontrou foi acrescentar aos modelos da Lógica de Primeira Ordem um conjunto de mundos possíveis. Na semântica relacional de Kripke, a sentença "necessário **p**" é verdadeira num mundo **w** quando **p** é verdadeira em todo mundo **w'** acessível a **w**. Já a sentença "possível **p**" é verdadeira num mundo **w** quando **p** é verdadeira em algum mundo **w'** acessível a **w**.

Assim, (6) será verdadeira se houver ao menos um mundo possível no modelo em que a sentença "Aristóteles foi discípulo de Platão" for falsa. Se em todos os mundos do modelo a sentença "Aristóteles foi discípulo de Platão" for verdadeira, teremos nesse modelo que (6) será falsa, pois não seria possível nesse modelo que Aristóteles não fosse discípulo de Platão.

Se entendermos as constantes individuais como contrapartida formal aos nomes próprios, torna-se evidente a relação estreita entre a Lógica Modal

[17] Isso não ocorre, porém, em todos os casos. Kripke chama atenção para o caso de Netuno, cujo nome próprio não foi dado por meio de um batismo inicial, mas como abreviação da descrição definida "o planeta a perturbar a órbita de Urano" (2012, p. 137).

de Primeira Ordem e a tese de que nomes próprios são designadores rígidos. A função de atribuição de um modelo atribui a uma constante individual sempre o mesmo objeto do domínio, independentemente do mundo possível. Assim, as constantes individuais são claramente designadores rígidos nesse sistema. O passo seguinte foi defender, tendo esses modelos em mente, que os nomes próprios na linguagem natural são as constantes individuais da linguagem formal.

Já uma descrição definida do tipo "o discípulo de Platão e mestre de Alexandre Magno" entendida à la Russell como "existe um único **x** tal que **x** foi discípulo de Platão e mestre de Alexandre Magno" pode ser verdadeira ou falsa, dependendo do mundo possível em que ela for interpretada. Num mundo possível em que ela fosse falsa, a descrição definida "o discípulo de Platão e mestre de Alexandre Magno" não se referiria a ninguém. Num mundo possível em que ela fosse verdadeira, a função de atribuição poderia atribuir a **x**, por exemplo, o indivíduo Espeusipo. Nesse mundo possível, a descrição "o discípulo de Platão e mestre de Alexandre Magno" se referiria a Espeusipo. Nesse sentido, descrições definidas não são rígidas, pois sua referência varia de mundo para mundo.

A tese de que as constantes individuais em sistemas modais devem ser entendidas como nomes próprios da linguagem natural foi claramente o que inspirou a filosofia da linguagem de Kripke.[18] Também não é difícil acomodar essa tese com a visão russeliana de que as descrições definidas da linguagem natural devem ser compreendidas como sentenças envolvendo existência e unicidade. Desse modo, está clara a relação intrínseca entre a filosofia da linguagem de Kripke e a Lógica Modal de Primeira Ordem.[19]

[18] Isso é admitido pelo próprio Kripke (2012, p. 34-44).

[19] No artigo *A Completeness Theorem in Modal Logic,* a Lógica Modal de Primeira Ordem tratada por Kripke é uma extensão de primeira ordem do sistema modal **S5** de Lewis e Langford**.** Seria temerário, porém, afirmarmos que a lógica da filosofia da linguagem de Kripke é apenas a extensão de primeira ordem de **S5** por algumas razões. Em seu artigo posterior *Semantical Analysis of Modal Logic I,* Kripke acrescenta aos modelos proposicionais de **S5** uma relação de acessibilidade entre os mundos possíveis (1963, p. 68), obtendo o resultado de completude para o sistema **S4** de Lewis e Langford (1963, p. 82). Em *Semantical Analysis of Modal Logic II*, Kripke enriquece ainda mais seus modelos, dividindo o conjunto de mundos possíveis em *normais* e *não-normais* (1966, p. 210) tendo como resultado a completude dos sistemas **S2** e **S3** de Lewis e Langford (1966, p. 214-217). Não há nenhuma razão para acreditar que Kripke considerava mais legítima do ponto de vista filosófico a extensão de primeira ordem de **S5** em detrimento às extensões de primeira ordem de **S4**, **S3** ou

Mais do que isso, somos forçados a admitir que a filosofia da linguagem de Kripke é fruto de sua semântica formal de mundos possíveis. Essas considerações nos obrigam a colocar Kripke, juntamente a Russell, no grupo de lógicos que entendem a linguagem humana como uma maquinaria complexa, e que cabe ao filósofo desvendar o seu funcionamento. Também está mais do que claro que, para Kripke, não há, como em Frege, duas lógicas. O que há é uma única lógica, que é a lógica da sua filosofia da linguagem.

Por fim, a Lógica Modal de Primeira Ordem de Kripke claramente respeita o princípio **PC** num sentido mais fraco: o valor de verdade de uma sentença complexa envolvendo modalidades como (6) é determinada pelo valor de verdade da sentença subordinada menos complexa no nosso mundo e em outros mundos possíveis. Nesse sistema, vale claramente tanto **P1** quanto **P2**. Mas **P2** vale apenas num sentido mais fraco: todos os designadores rígidos se referem a um único indivíduo existente no nosso mundo atual, mas que poderia não existir em outros mundos possíveis.[20] Tendo essas ponderações em mente, podemos inferir que a Lógica Modal de Primeira Ordem é uma lógica do tipo clássico. Foi apenas para fins esquemáticos que, na introdução desse artigo, denominamos a lógica de Kripke de simplesmente Lógica Modal Clássica, por mais vaga e simplista que seja essa denominação.

S2. De fato**,** Kripke não parece se comprometer filosoficamente com nenhum tipo de acessibilidade entre os mundos possíveis.

[20] Está claro em *O Nomear e a Necessidade* que os objetos referidos pelos designadores rígidos não são necessariamente existentes. Isso porque Kripke denominava de designador fortemente rígido aqueles nomes próprios e descrições definidas que se referem a um único indivíduo necessariamente existente, ou seja, existente em todos os mundos possíveis, não apenas no nosso mundo atual. Esse seria o caso, por exemplo, do numeral "1" e da descrição definida "o sucessor de 0": ambos os termos se referem a um único indivíduo, o número 1, que existe em todos os mundos possíveis (KRIPKE, 2012, p. 99). Porém, no mesmo texto não está tão claro que um designador rígido deva necessariamente se referir a um indivíduo existente no nosso mundo atual ou, ao menos, em algum mundo possível. Kripke diz simplesmente que "uma expressão é um designador *rígido* se ela designar o mesmo objeto em todos os mundos possíveis" (KRIPKE, 2012, p. 99). Em sua introdução para a tradução portuguesa, Santos defende que "a questão de saber se um nome continua a designar o mesmo objeto nos mundos possíveis em que esse objeto não existe é controversa" e emenda que o próprio Kripke parece oscilar em outras obras a respeito dessa questão (KRIPKE, 2012, p. 23).

5. Kaplan e os Indexicais

Russell respondeu às objeções de Strawson a sua teoria das descrições definidas de maneira leviana e impaciente. Para ele, Strawson se ateve demais à palavra "atual". Segundo Russell, esse é um tipo de palavra "egocêntrica". Palavras egocêntricas podem se referir a objetos distintos em situações diferentes e, portanto, são intrinsecamente ambíguas. Mas a ambiguidade pode ser facilmente eliminada se substituirmos a sentença (5) pela proposição

(5p) O rei da França no ano de 1905 é sábio

que substitui, em (5), o termo "atual" pelo ano em que o artigo *Da Denotação* foi publicado (RUSSELL, 1957, p. 385).[21]

Aqui, fica clara a técnica de Russell: as chamadas palavras egocêntricas devem ser eliminadas, de modo que toda sentença em que elas ocorrem deve ser substituída, na análise lógica, por uma sentença equivalente em significado que elimine essas palavras. Isso ocorre não apenas com "atual", mas também com palavras como "eu", "aqui", "agora", etc.

Enquanto Russell classifica essas palavras como "egocêntricas", Kaplan em seu artigo *Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals* as denomina de *indexicais*. De acordo com Kaplan, indexicais englobam: pronomes como "eu", "meu", "ele", "seu", "ela", "sua"; pronomes demonstrativos como "esse", "aquele"; advérbios como "aqui", "agora", "ontem", "amanhã"; adjetivos como "atual", "presente", dentre outros (1989, p. 489).

Para Kaplan, os indexicais devem ser divididos em dois grupos. O primeiro grupo é formado por indexicais em que o referente é fixado por meio de uma apresentação, normalmente visual, do objeto, ao apontarmos para ele. Esse é o caso dos indexicais "este" e "aquele". Tais indexicais são chamados

[21] Outro ponto enfatizado por Russell em sua resposta a Strawson: enquanto Strawson está propondo fazer filosofia com a linguagem cotidiana, Russell está defendendo que uma linguagem formal, científica é necessária para fazer filosofia do mesmo modo que é necessário em outros campos, como na física e na química (1957, p. 387). Esse é mais um argumento a favor da tese defendida na Seção 2.1 deste artigo: para Russell, não há distinção entre lógica e filosofia da linguagem.

de *demonstrativos*. Já indexicais como "eu", "agora" e "amanhã" são indexicais *puros*, pois não necessitam de um ato de demonstração para fixar o referente. Aqui, é o contexto que determina a referência (KAPLAN, 1989, p. 491-492).

Do ponto de vista semântico, Kaplan não propõe uma visão alternativa a de Kripke, e sim complementar. Vimos que, na semântica kripkeana, a função que determina o valor de verdade depende de uma proposição e de um mundo possível. Visando um comprometimento metafísico menor, Kaplan denomina os mundos possíveis de Kripke de "circunstâncias possíveis de valoração" (1989, p. 507).

Mas Kaplan faz uma complementação à teoria de modalidade de Kripke que não é simplesmente apresentá-la sobre uma nova roupagem. Além da função de verdade que associa a cada proposição em cada mundo um certo valor de verdade, há uma outra função para Kaplan. Essa outra função associa a cada sentença num certo contexto uma determinada proposição. Para compreendermos essa distinção, tomemos o seguinte exemplo:

(8) Eu poderia não ter sido discípulo de Platão

Imaginemos um contexto em que a sentença (8) seja dita por Aristóteles. Aqui, está claro que o significado de (8) nesse contexto é a mesma proposição expressa por (6), a saber, que Aristóteles poderia não ter sido discípulo de Platão. Dado esse conteúdo proposicional fixado pelo contexto, podemos decidir se há um mundo possível em que Aristóteles não é discípulo de Platão para, aí sim, determinar o valor de verdade da sentença (8) no contexto em que Aristóteles a pronuncia.

Agora, consideremos um outro contexto em que a mesma sentença (8) foi pronunciada por Espeusipo em vez de pronunciada por Aristóteles. Aqui, temos que considerar as circunstâncias possíveis de valoração de outra proposição, a de que Espeusipo poderia não ter sido discípulo de Platão.

O que concluímos daqui é que, para determinar se a sentença (8) é verdadeira ou falsa, é preciso, primeiramente, determinar que proposição é

expressa por (8) num certo contexto. Em seguida, deve-se levar em conta as circunstâncias possíveis de valoração – os mundos possíveis – para, aí sim, determinarmos o valor de verdade de (8).

Vimos que as circunstâncias possíveis de valoração são, no vocabulário kaplaniano, os mundos possíveis de Kripke. Mas os contextos, por sua vez, determinam o que Kaplan denomina de *conteúdo* de uma sentença, que é o que normalmente chamamos de proposição. De acordo com Kaplan, quando Strawson (STRAWSON, 1950, p. 326) diz que usamos a sentença (5) para fazer asserções (*statements*) distintas, ele está afirmando, em outras palavras, que há conteúdos proposicionais distintos da mesma sentença em contextos distintos (1989, p. 501).[22]

Cabe ressaltar que a noção de conteúdo não cabe apenas a uma sentença, mas também às partes de uma sentença (KAPLAN, 1989, p. 501). Desse modo, o que faz a sentença (5) ser verdadeira ou falsa se enunciada em momentos distintos é a referência do indexical "atual". No contexto do reinado de Luís XIV, o termo singular "o atual rei da França" se refere ao indivíduo Luís XIV, de modo que poderíamos aceitar o conteúdo de (5) como verdadeiro. Já no contexto do reinado de Luís XV, o mesmo termo singular "o atual rei da França" se refere ao indivíduo Luís XV, de modo que poderíamos tomar o conteúdo de (5) com sendo falso.

5.1. A "lógica" de Kaplan

Kaplan defende que o significado (*meaning*) tem duas dimensões: um *caráter* e um *conteúdo*. O caráter é uma função de contextos para conteúdos. O conteúdo é uma função de circunstâncias possíveis para extensões. Em seguida, Kaplan diz explicitamente que ambas as funções respeitam **PC** (KAPLAN, 1989, p. 505-507)

[22] Kaplan (1989, p. 501) também chama de conteúdo de uma proposição aquilo que Frege (FREGE, 2009, p. 239) denominava de sentido de uma proposição, que é o pensamento. O autor, porém, pondera que o conceito fregeano de sentido de uma sentença confunde a noção epistemológica de diferença de cognição (por que sentenças do tipo **t = t** difere em conteúdo cognitivo de sentenças do tipo **t = t'**, em que **t** e **t'** são termos singulares) com a noção semântica de conteúdo de uma sentença. Com relação ao papel dos indexicais para determinar o sentido de uma sentença, diz Frege: "o mesmo enunciado que encerre a palavra 'eu' expressará, quando proferido por diferentes pessoas, diferentes pensamentos, alguns dos quais poderão ser verdadeiros, e outros, falsos" (FREGE, 2002, p. 20).

A semântica formal de Kaplan consistia, a princípio, em enriquecer os modelos kripkeanos de modo que um contexto fosse determinado por uma quádrupla **(w, x, p, t)** em que **w** é um mundo possível, **x** é um agente, **p** é um ponto no espaço e **t** é um instante no tempo. Mas tal enriquecimento não é o suficiente para manter distinções intuitivas em sentenças envolvendo indexicais. Tomemos o seguinte exemplo de Kaplan:

(9) Eu estou aqui agora.

Dado um certo contexto **(w, x, p, t),** o conteúdo da sentença (9) é o mesmo que o da sentença:

(9p) David Kaplan está em Portland em 26 de Março de 1977.

Mas enquanto a sentença (9) parece ser universalmente verdadeira, (9p) é contingentemente verdadeira (KAPLAN, 1989, p. 508-509). A solução encontrada por Kaplan para esse problema foi primeiramente apresentada em seu artigo *On the Logic of Demonstratives*.

Nesse artigo, Kaplan defende que os predicados e as funções devem ser duplamente indexados. Assim, enquanto na Linguagem de Primeira Ordem com Igualdade temos predicados n-ários e funções n-ária,[23] na Linguagem dos Demonstrativos teremos predicados m-n-ários e funções m-n-árias. (KAPLAN, 1979, p. 87)

Semanticamente, um predicado **P** é interpretado como uma função em que, dado um certo mundo **w** e um certo instante de tempo **t**, temos o conjunto dos indivíduos que têm a propriedade expressa por **P** naquele mundo **w** e naquele instante de tempo **t**. Já as funções se dividem em *individuais* e *posicionais*. No caso das funções individuais, elas são interpretadas como funções que levam indivíduos e posições em indivíduos. No caso das funções posicio-

[23] Por exemplo, o predicado "**x** é mortal" é unário, enquanto o predicado "**x** é discípulo de y" é binário; a função "a mãe de **x**" é unária, enquanto a função "o filho mais velho de **x** e **y**" é binária.

nais, elas são interpretadas como funções que levam indivíduos e posições em posições. (KAPLAN, 1979, p. 88-89)

Para ilustrar tal semântica, retomemos o problema da referência do termo singular "o atual rei da França". Na semântica formal acima exposta, estamos diante de uma função individual **o_rei_da_França(x)**. Dado um contexto **(w, x, p, t)**, a função **o_rei_da_França(x)** aponta o indivíduo que é rei da França naquele contexto, caso haja um. Caso não haja nenhum rei na França naquele contexto, a função retorna como valor o símbolo arbitrário †.

Daqui, podemos inferir que a filosofia da linguagem de Kaplan é um reflexo bastante fiel da sua linguagem formal dos demonstrativos e sua semântica formal dos demonstrativos. Aqui, estamos claramente diante de um filósofo que vê a linguagem humana como uma maquinaria complexa, de modo que sua semântica formal captaria o funcionamento da linguagem natural. Podemos, nesse sentido, colocar Kaplan no mesmo grupo de Russell e Kripke, como fizemos na introdução deste artigo.

Mas o leitor atento pode ter se dado conta que, em vez de colocarmos a Lógica dos Demonstrativos como a lógica de Kaplan na tabela da introdução, colocamos a Lógica Híbrida em seu lugar. De fato, a Lógica dos Demonstrativos contém uma semântica formal e uma linguagem formal que, como vimos, é bem fidedigna à filosofia da linguagem de Kaplan. Mas em nenhum de seus trabalhos Kaplan apresenta um sistema de prova para essa semântica (com regras e axiomas) e, muito menos, prova algo semelhante ao Teorema de Completude que Kripke provou para a Lógica Modal de Primeira Ordem.

O ponto aqui é sabermos em que medida podemos considerar como lógica uma linguagem formal e semântica formal sem um conjunto de regras e axiomas, ou seja, sem um sistema dedutivo completo com relação a essa semântica. Em *Filosofia das Lógicas*, a filósofa Haack expõe que, de acordo com Kneale e outros lógicos, para um sistema formal ser considerado uma lógica, ele precisa ser completo com relação a uma certa semântica (KNEALE, 1961, p. 258-259 apud HAACK, p. 32). A filósofa, por sua vez, argumenta que esse critério "formal" de Kneale não é adequado, visto que a incompletu-

de da teoria de conjuntos mostraria que o conceito de "pertinência" não pode ser caracterizado por um conjunto de axiomas e regras.

Ainda que o critério de Kneale possa não ser unânime para decidir se um sistema de prova ou uma semântica formal é ou não uma lógica, usarei aqui tal critério para distinguir lógica formal em sentido lato de lógica formal em sentido estrito. Em sentido lato, lógica formal pode ser tanto sinônimo de uma certa semântica formal quanto de um sistema de prova. Em sentido estrito, lógica formal consiste de uma semântica formal, um sistema de prova e um Teorema de Completude, que garante que são sempre verdadeiros semanticamente os teoremas provados no sistema, e apenas tais teoremas.

Em sentido lato, a lógica dos demonstrativos de Kaplan é claramente uma lógica dos indexicais. Ou seja, tal "lógica" consiste (apenas) de uma semântica formal que procura apreender a filosofia da linguagem de Kaplan. Mas seria possível uma lógica dos indexicais em sentido estrito? Acredito que esse é justamente o caso da lógica híbrida, como argumentarei a seguir.

5.2. Lógica Híbrida: a Lógica dos Indexicais em Sentido Estrito

Numa lógica modal, a função de verdade atribui um valor de verdade a uma dada proposição **p** e um mundo possível **w**. Isso significa que a verdade é relativa a cada mundo específico **w**. Para obter uma linguagem modal formal, normalmente acrescenta-se a uma linguagem formal apenas o operador modal de necessidade, definindo-se, por meio da negação e da necessidade, o operador de possibilidade.[24] (BRAÜNER, 2020)

Como exemplo do problema de expressividade da lógica modal, retomemos a sentença (5). A princípio, podemos tomar os mundos possíveis como instantes de tempo **t**, **t'**, **t"** e assim por diante.[25] Assim, a sentença (5) poderia ser representada modalmente como:

[24] Essa é apenas uma maneira de definir uma linguagem modal formal. Outra maneira de definir uma linguagem modal formal equivalente é tomar o operador de possibilidade como primitivo e definir o operador de necessidade. Assim, dada uma proposição "**p**", a operação "é necessário **p**" seria equivalente, por definição, a "não é possível não-**p**".

[25] Nessa interpretação, os operadores modais são compreendidos de modo distinto. Aqui, "**p** é necessário" seria o mesmo que "em todos os instantes de tempo", ou simplesmente, "sempre". Já "**p** é possível" seria o mesmo que "em algum instante de tempo", ou simplesmente, "alguma vez".

(5t) O rei da França no instante **t** é sábio.

O ponto aqui é que a linguagem modal pode apenas expressar que o rei da França é sábio em todos os instantes de tempo ou em ao menos um instante de tempo indeterminado. Não é possível, porém, expressar nessa linguagem um certo instante de tempo específico, que seria, por exemplo, o ano de 1905.

Foi com essa intenção que Gargov e Goranko propuseram, em seu artigo *Modal Logic with Names,* enriquecer a linguagem modal com aquilo que denominaram *constantes modais* ou simplesmente *nomes* (*names*). Na linguagem formal proposta pelos autores, essas constantes modais são consideradas fórmulas atômicas, semelhantemente às variáveis proposicionais. Além disso, acrescenta-se um *operador modal universal* à linguagem. Semanticamente, interpreta-se as contantes modal do seguinte modo: acrescenta-se aos modelos de Kripke uma função que, a cada constante modal, atribui-se um mundo específico do conjunto de mundos possíveis. Cabe ressaltar que há uma diferença importante entre constantes modais e variáveis proposicionais: enquanto o valor semântico de uma variável proposicional pode ser entendido como um conjunto de mundos possíveis em que a proposição é verdadeira, o valor semântico de uma constante modal é um único mundo possível. Já o operador modal de necessidade universal é interpretado do ponto de vista semântico analogamente a um quantificador universal que, em vez de percorrer o domínio de objetos, percorre o domínio de mundos possíveis.[26] (GARGOV e GORANKO, 1993, p. 610-611)

Mas será que essa linguagem híbrida consegue formalizar o indexical "atual" em (5)? Para isso, basta tomar uma constante modal **a** como contrapartida formal de um instante de tempo específico, por exemplo o ano de

[26] O leitor pode estar se perguntando qual seria a diferença entre o operador de necessidade kripkeano e o operador de necessidade universal, proposto por Gargov e Goranko. Vimo que operador de necessidade de Kripke é interpretado de modo que "necessário **p**" é verdadeiro num mundo **w** quando **p** é verdadeiro em todos os mundos possíveis **w'** acessíveis a **w.** Já "necessário universalmente **p**" é verdadeiro quando **p** é verdadeiro em todos os mundos possíveis **w',** independentemente de serem ou não acessíveis a **w**. Apenas no sistema **S5**, em que a relação de acessibilidade entre os mundos é universal, que os operadores de necessidade e de necessidade universal se colapsam**.**

1905. Desse modo, conseguimos expressar nessa linguagem formal que o rei da França é sábio no ano de 1905, o que não conseguiríamos fazer na linguagem modal de Kripke.

A diferença principal entre essa abordagem e a "lógica" de Kaplan é que estamos diante de uma lógica muito mais madura. Na Seção 5 do artigo, os autores apresentam um sistema de prova com regras e axiomas (GARGOV e GORANKO, 1993, p. 622), demonstrando na Subseção 5.3 a completude desse sistema com relação à semântica anteriormente apresentada (GARGOV e GORANKO, 1993, p. 625-628).

Evidentemente, o sistema de Gargov e Goranko é muito mais rudimentar do que a semântica de Kaplan, pois a quádrupla **(w, x, p, t)** foi reduzida a um único mundo **w**. Mas não é nada muito complexo obter axiomas e uma semântica adequada para um sistema em que há quatro tipos de constantes modais: para mundos possíveis, para agentes, para posições e para instantes.

Outro ponto a enfatizar quanto à pobreza expressiva do sistema de Gargov e Goranko é que ele é proposicional enquanto que a filosofia da linguagem de Kaplan é claramente de primeira ordem. Numa linguagem proposicional, não há variáveis para indivíduos, apenas para proposições enquanto que, como vimos, numa Linguagem de Primeira Ordem há variáveis para termos gerais e para termos singulares. Assim, o problema do termo singular "o rei da França" se referir a indivíduos distintos em instantes de tempo distintos sequer pode ser colocado no sistema de Gargov e Goranko.

Mas não há nenhuma dificuldade adicional em estender os resultados de Gargov e Goranko para uma Lógica Híbrida de Primeira Ordem. Na verdade, Braüner foi o primeiro a propor uma Linguagem Híbrida de Primeira Ordem. Essa linguagem é uma expansão da Linguagem Modal de Primeira Ordem em que acrescentam-se: (i) variáveis para constantes modais; (ii) variáveis para descrições definidas; (iii) um operador unário para "o único indivíduo **i** em **a**", tal que **a** é um nome para o certo mundo **w**. Além disso, nessa linguagem há um novo quantificador, denominado *aglutinante* (*binder*). Se **a** é um nome para o mundo **w**, podemos expressar por meio desse quantificador

algo como "a extensão do predicado **P** no mundo de nome **a**" (BRAÜNER, 2011, p. 129-131). Na mesma obra, Braüner propõe como sistema de prova um sistema de Dedução Natural para Lógica Híbrida de Primeira Ordem (2011, p. 139-140), demonstrando o Teorema de Completude com relação a essa lógica (2011, p. 143-149).

6. Considerações Finais

A Lógica Híbrida de Primeira Ordem de Braüner é construída por meio de uma linguagem formal bastante complexa. Essa linguagem captura conceitos modais como *necessário* e *possível,* bem como a distinção entre nomes próprios e descrições definidas. Além disso, tal linguagem é rica o suficiente para expressar algo do tipo "no nosso mundo atual". Assim, um termo singular como "o atual rei da França" pode ser expresso nessa linguagem como "o indivíduo denotado pela descrição definida 'o único rei da França' no mundo denotado por **a**".

Porém, vimos na Seção 3 que, para Strawson, há mundos em que "o atual rei da França" não denota ninguém. Para isso, precisamos de uma lógica livre, ou seja, uma lógica que não respeita **P2**. Dito de outro modo, precisamos uma lógica em que nem todo termo singular se refere a um único objeto.

Vimos também, na mesma Seção 3, que se numa sentença a expressão "o atual rei da França" não tem referência, então, para Strawson, a sentença como um todo não será nem verdadeira nem falsa. Desse modo, a lógica da filosofia da linguagem de Strawson parece, como a de Frege, não respeitar **P1** mas respeitar **PC**. Além disso, vimos na Subseção 1.1 que a lógica livre de Lehmann, denominada NINO, é justamente uma lógica que não respeita **P1** e **P2**, mas respeita **PC.** O problema, como apontamos na Seção 3.1, é como interpretar formalmente em NINO o termo "atual". Mas é justamente aqui que entra em cena a Lógica Híbrida de Primeira Ordem de Braüner.

Vejamos então como interpretar a filosofia da linguagem de Strawson numa combinação entre a lógica de Lehmann e a lógica de Braüner. Na lógica em questão, a sentença (5) poderia ser expressa como:

(5h) "O rei da França no mundo **a** é sábio".

Seja **w** nosso mundo atual tal que o nome de **w** é **a**. Seja **P** o predicado unário "**x** é sábio" e **i** o termo individual para "o rei da França". Como não há um rei da França, o termo **i** não se refere a ninguém no mundo **w**. Portanto, a sentença (5h) não será nem verdadeira nem falsa.

Agora, suponhamos que **a** seja um nome de um outro mundo diferente do nosso, o mundo **w'**. Nesse mundo, o termo **i** se refere a Luís XIV tal que o indivíduo Luís XIV está na extensão no predicado "**x** é um sábio". Logo, nesse mundo **w',** a sentença (5h) será verdeira.

Por fim, consideremos que **a** é o nome de um terceiro mundo diferente dos outros dois, o mundo **w"**. Nesse mundo, o termo **i** se refere a Luís XV tal que o indivíduo Luís XV não está na extensão no predicado "**x** é um sábio". Logo, nesse mundo **w"** a sentença (5h) será falsa.

Por essas razões, acredito que uma combinação da lógica híbrida de Braüner com a lógica NINO de Lehmann seria muito mais próxima da filosofia da lingugem de Strawson do que as duas opções apresentadas na Subseção 3.1: a de Caorsi e a de van Fraassen.

6.1 Por uma Lógica Híbrida Livre de Quatros Valores

Ainda que uma combinação da lógica híbrida de Braüner com a lógica NINO possa resolver em grande medida os problemas concernentes aos termos singulares que foram abordados nesse artigo, ela não dá conta da seguinte questão. Retomemos as sentenças abaixo:

(2) Ulisses desembarcou em Ítaca;

(3) Ulisses não desembarcou em Ítaca.

Como apontado por Orlando em seu artigo *Fictional Names without Fictional Objects,* uma perspectiva que oferece um modelo para a sentença (2) ser verdadeira – como na lógica livre positiva – é tão anti-intuitiva como outra

que infere que a sentença (3) é verdadeira – como na lógica livre negativa (ORLANDO, 2008, p. 118). Poderíamos acrescentar ainda que uma perspectiva que infere que é consistente sustentar, ao mesmo tempo, (2) e (3) também seria bastante anti-intuitiva, que é justamente o que diz tanto a lógica livre de Lehmann quanto a lógica livre de Bencivenga.

Para superar esse dilema, proponho distinguir dois níveis de verdade: verdades ficcionais e verdades reais. Acredito que tal lógica da filosofia da linguagem consiste numa lógica livre de quatro valores. Considerando que o termo ficcional "Ulisses" não tem referência, a sentença (2) como um todo deve ser ficcionalmente verdadeira e a sentença (3) deve ser ficcionalmente falsa. Podemos generalizar essa ideia para quaisquer nomes ficcionais, de modo que teremos quatro valores: verdadeiro realmente, verdadeiro ficcionalmente, falso ficcionalmente e falso realmente.

É digno de nota que essa lógica, diferentemente das lógicas livres existentes na literatura, respeitaria **P1** de um modo mais fraco, a saber: toda sentença seria verdadeira (realmente ou ficcionalmente) ou falsa (realmente ou ficcionalmente), não podendo ser atribuído a cada sentença mais de um valor de verdade ou nenhum. Por outro lado, essa lógica respeitaria a condição de Strawson também num sentido mais fraco: em contextos ficcionais, a verdade real ou a falsidade real de uma sentença não seria colocada, ainda que pudesse ser colocada a questão da verdade ou falsidade ficcional da sentença.

Por fim, combinando uma lógica livre de quatro valores com a lógica híbrida de Broüner, podemos ter uma linguagem rica o suficiente para, em grande medida, acomodar também as modalidades de Kripke e os indexicais de Kaplan. Por essas razões, acredito que a filosofia da linguagem que emergisse dessa lógica seria capaz de propor um modelo que abarcasse as considerações de Frege, Russell, Strawson, Kripke e Kaplan sobre os termos singulares. Além disso, tal lógica passaria incólume às críticas apontadas por Orlando às lógicas livres existentes na literatura.

Não quero dizer que minha proposta de uma lógica híbrida livre de quatro valores seria a solução definitiva para o problema dos termos singulares

na linguagem natural. Simplesmente porque – parodiando Strawson – a lógica exata dos termos singulares na linguagem natural simplesmente não existe.

Agradecimentos

Essa pesquisa foi possível graças ao financiamento institucional da UFFS (Universidade Federal da Fronteira Sul) por meio do Projeto PES-2019-0591 intitulado "Nomes próprios e os limites das teorias referenciais da linguagem".

Agradeço a leitura atenta do parecerista anônimo da revista, o que permitiu que a versão final do artigo tivesse uma versão muito mais apresentável que a original. Também agradeço a leitura cuidadosa dos discentes Deródio Felipe Perondi Meotti (UFFS) e Nathalia Rodrigues da Costa (Unicamp) que, mesmo não sendo estudiosos em lógica, revisaram o artigo minuciosamente. Agradeço, finalmente, os demais colegas, docentes e discentes, membros do Grupo de Pesquisa em Lógica, Linguagem e Conhecimento da UFFS que tiveram o trabalho de ler versões prévias desse texto, com comentários e observações gerais.

Referências

ARISTÓTELES, *"De Interpretatione"*, trad. J. L. Ackrill, em: ROSS, D. (ed.), **The Complete Works of Aristotle**, Princeton: Princenton University Press,1991.

_____, *"Prior Analytics"*, trad. A. J. Jenkinson, em: ROSS, D. (ed.), **The Complete Works of Aristotle**, Princeton: Princenton University Press,1991.

BENCIVENGA, E , "Free Logics," em: GABBAY, D. e GUENTHNER, F. (eds.), **Handbook of Philosophical Logic**, vol. III: Alternatives to Classical Logic, Dordrecht: D. Reidel, 1986.

BRAÜNER, T., **Hybrid Logic and its Proof-Theory,** em: "Applied Logic Series": Volume 37, Dordrecht-Heidelberg-Berlin-New York: Springer, 2011.

_____, "Hybrid Logic", em: **The Stanford Encyclopedia of Philosophy**, ZALTA, E. N. (ed.) Disponível em: https://plato.stanford.edu/entries/logic-hybrid/ ,11 de julho de 2020.

CAORSI, C. E. "Referência e termos singulares", trad. Itamar Luís Gelain, em: CONTE, J. (ed.), **Ensaios Sobre a Filosofia de Strawson**, Florianópolis: UFSC, 2015.

DONNELLAN, K., "Reference and Definite Descriptions", em: **Philosophical Review,** v. 75, 1966, p. 281–304.

FREGE, G., "Digressões sobre Sentido e Referência" em: **Lógica e Filosofia da Linguagem.** 2. ed. revista e ampliada. Seleção e trad. de Paulo Alcoforado. São Paulo: EDUSP, 2009.

_____, **Fundamentos da Aritmética**, em: "Coleção Os Pensadores", vol. XXVI, trad. Luiz Henrique dos Santos, São Paulo: Abril Cultural, 1974.

_____, "O Pensamento. Uma Investigação Lógica", em: **Investigações Lógicas**. Org., trad. e notas de Paulo Alcoforado, Porto Alegre, EDIPUCRS, 2002.

_____, **Os Primeiros Escritos Lógicos de Gottlob Frege**, tradução de P. Alcoforado,A. Duarte e G. Wyllie,. São Paulo: Instituto Brasileiro de Filosofia e Ciência Raimundo Lulio, 2012.

_____, "Sentido e Referência" em: **Lógica e Filosofia da Linguagem**. 2. ed. revista e ampliada. Seleção e trad. de Paulo Alcoforado. São Paulo: EDUSP, 2009.

_____, "Sobre a Justificação Científica de uma Conceitografia" em: **Lógica e Filosofia da Linguagem**. 2. ed. revista e ampliada. Seleção e trad. de Paulo Alcoforado. São Paulo: EDUSP, 2009.

GARGOV, G., GORANKO, V., "Modal Logic with Names", **Journal of Philosophical Logic**, vol. 22, 1993, p. 607-636.

HAACK, S. **Filosofia das Lógicas**. Tradução de César Augusto Mortari. São Paulo: Unesp, 2002.

HEGENBERG, L. **Lógica**, Rio de Janeiro, RJ: Editora Forense Universitária, 2012.

KAPLAN, D. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals" em: PERRY A., J., WETTSTEIN, H., **Themes from Kaplan**, Oxford: Oxford University Press, 1989.

_____, "On the Logic of Demonstratives", em: **Journal of Philosophical Logic**, vol. 8, n. 1, janeiro de 1979, p. 81-98.

KNEALE, W.C., "The Province of Logic", em: **Contemporary British Philosophy,** Lewis (ed.), série 3, 2ª edição, Allen and Unwin, 1961.

KRIPKE, S., **O Nomear e a Necessidade**, trad. Ricardo Santos e Teresa Filipe, Lisboa: Gradiva, 2012.

_____, "A Completeness Theorem in Modal Logic", em: **The Journal of Symbolic Logic,** v. 24, n. 1, março de 1959., p. 1-14.

_____, "Semantical Analysis of Modal Logic I. Normal Modal Propositional Calculi", em: **Zeitschrift für Mathematische Logik und Grundlagen der Mathematik**, v. 9, n. 5–6: 67–96, 1963.

_____, "Semantical Analysis of Modal Logic II. Non-normal Modal Propositional Calculi", em: ADDISON, J.W.; HENKIN, L.; TARSKI, A. (eds.), **Symposium on the Theory of Models,** Amsterdam: North-Holland, 1965.

LEHMANN, S., "Strict Fregean Free Logic", em: **Journal of Philosophical Logic**, v. 23 n. 3: 307–336, 1994.

MORRIS, M, **An introduction to Philosophy of Language**, Cambridge: Cambridge University Press, New York, NY, 2007. 326 p.

MORTARI, C. A., **Introdução à Lógica**. 2 ed., São Paulo: Unesp, 2016. 525 p.

NOLT, J., "Free Logic", em: **The Stanford Encyclopedia of Philosophy**, ZALTA, E. N. (ed.) Disponível em: https://plato.stanford.edu/archives/fall2018/entries/logic-free/, 11 de julho de 2020.

ORLANDO, L., "Fictional Names without Fictional Objects", em: **Revista Hispanoamericana de Filosofía**, vol. 40, n. 120, p. 111-127.

RUSSELL, B., **A Filosofia do Atomismo Lógico**, em: "Coleção Os Pensadores", vol. XLII, trad. Pablo Rubén Mariconda, São Paulo: Abril Cultural, 1974a.

_____, **Da Denotação**, em: "Coleção Os Pensadores", vol. XLII, trad. Pablo Rubén Mariconda, São Paulo: Abril Cultural, 1974b

_____, **Introdução à Filosofia Matemática**, trad. Maria Luíza X. de A. Borges, Rio de Janeiro : Jorge Jahar, 1996

_____, "Mr. Strawson on Referring", em: **Mind**, vol. 66, n. 263, 1957, p. 385-389.

_____, "On Denoting", em: **Mind**, vol. 14, 1905, p. 479–93.

_____, **Os Problemas da Filosofia**, trad. António Sérgio, Coimbra: Almedina, 2001.

STRAWSON, P. F., "On Referring",em: **Mind**, v. 59, 1950, p. 320–44.

_____, **Sobre o Referir**, em "Coleção Os Pensadores", vol. LII, trad. Balthazar Barbosa Filho, São Paulo: Abril Cultural, 1974.

_____, "Réplica a Carlos C. Caorsi" em: CAORSI, C. C (ed.), **Ensaios Sobre a Filosofia de Strawson** com réplicas de P. F. Strawson, trad. Itamar Luís Gelain e Jaimir Conte, Ijuí: Ed. Unijuí, 2015.

SZABÓ, Z. G., "Compositionality", em: **The Stanford Encyclopedia of Philosophy**, ZALTA, E. N. (ed.). Disponível em: https://plato.stanford.edu/archives/sum2017/entries/compositionality/, 11 de julho de 2020.

van FRAASSEN, B. C., "Singular Terms, Truth Value Gaps and Free Logic," em: **Journal of Philosophy,** v. 63, n. 17, 1966, p. 481–95.

SENTIDO E VERDADE: REVISITANDO O PROBLEMA DOS ENUNCIADOS EXISTENCIAIS NEGATIVOS

Anderson Luis Nakano[1]

RESUMO

Este artigo revisita o clássico problema dos nomes próprios vazios, ou ainda, o problema dos enunciados existenciais singulares negativos. Partindo de um parentesco deste problema com o problema (ou paradoxo) do falso, o artigo mostra, em um primeiro momento, como Aristóteles introduz uma distinção entre "nome" e "declaração" com o intuito de separar as condições de sentido das condições de verdade de um enunciado, abrindo assim a possibilidade para que um discurso seja falso sem ser, por isso, destituído de sentido. Em seguida, o artigo mostra como a ideia de que o sentido é anterior à verdade é radicalizada no *Tractatus* de Wittgenstein, radicalização que tem, como uma de suas consequências, a necessidade de se distinguir entre nomes próprios ordinários (que serão tratados como equivalentes a descrições) e nomes próprios logicamente genuínos, para os quais sequer se coloca a questão da existência ou não-existência. Em um terceiro momento, a atenção se volta à obra de Kripke a fim de mostrar como este, ao negar que os nomes próprios da linguagem ordinária sejam equivalentes a descrições, vai chegar, em sua análise dos enunciados existenciais negativos, à recusa daquilo que Aristóteles e Wittgenstein punham como pressuposto, a saber, a anterioridade do sentido de um enunciado em relação à sua verdade ou falsidade. A partir disso, algumas conclusões um tanto quanto paradoxais são extraídas da análise de Kripke para estes enunciados. Por fim, o artigo busca fornecer uma via para entendê-las por meio de uma comparação, ainda que bastante breve, desses enunciados com aquilo que Wittgenstein chama, na sua última obra, de proposições fulcrais (*hinge propositions*).

Palavras-chave: Nomes Vazios. Enunciados Existenciais Negativos. Wittgenstein. Kripke. Proposições fulcrais.

ABSTRACT

In this paper, I revisit the classic problem of empty proper names or, more precisely, the problem of negative singular existential statements. I start by establishing a relationship between this problem and the problem (or paradox) of how false statements are possible. I then show, in a first moment,

[1] Doutor em Filosofia pela Universidade Federal de São Carlos (UFSCar) e professor da Pontifícia Universidade Católica de São Paulo (PUC-SP).
E-mail: alnakano@pucsp.br. ORCID: 0000-0002-2221-2333.

how Aristotle introduces a distinction between "name" and "declaration" in order to separate the meaning conditions from the truth conditions of a statement, thus opening the possibility for a discourse to be false without being meaningless. Afterwards, I show how the idea that meaning is prior to truth is radicalized in Wittgenstein's *Tractatus*. One of the consequences of this radicalization is the distinction between ordinary proper names (which are treated as equivalent to descriptions) and logically genuine proper names, for which the question of existence or non-existence does not even arise. In a third moment, I turn to Kripke's work to show how he, in denying that the proper names of ordinary language are equivalent to descriptions, is obliged, in his analysis of negative existential statements, to refuse what Aristotle and Wittgenstein presupposed, namely, that the meaning of a statement is prior to its truth or falsity. I then draw some seemingly paradoxical conclusions from Kripke's analysis of these statements. Finally, I provide a way to understand these conclusions by means of a brief comparison between these statements and what Wittgenstein calls, in his last work, hinge propositions.

Keywords: Empty Names. Negative Existential Statements. Wittgenstein. Kripke. Hinge Propositions.

1. Introdução

Neste artigo, revisitarei o clássico problema dos nomes próprios vazios, ou ainda, o problema dos enunciados existenciais negativos como, p. ex.: "Sherlock Holmes não existe" ou "Vulcano não existe"[2]. *Grosso modo*, o problema consiste em que, ao afirmar a respeito de algo a não-existência, estaríamos aparentemente pressupondo que este algo existe pois, do contrário, do que é que estaríamos negando a existência?[3] Tudo se passa como se faltasse, em um enunciado desse tipo, o objeto do discurso, o "quê" de que falamos sobre. Este problema mantém certo parentesco com o paradoxo do discurso falso, cuja conclusão é que não pode haver enunciado falso, precisamente por faltar, a este, o seu objeto (segundo o paradoxo, isso supostamente condenaria o enunciado a ser apenas uma elocução sem relação com a realidade, deixando de ser propriamente um enunciado). Com efeito, se seguirmos a sugestão de

[2] Trataremos, em particular, dos enunciados existenciais singulares, em que um nome próprio aparece como sujeito gramatical da sentença, em contraste com enunciados como "Quimeras não existem", por exemplo.

[3] Em *Some Main Problems of Philosophy*, G. E. Moore expõe o problema do seguinte modo: "It seems as if there must be such a thing, merely in order that it may have the property of not being. It seems, therefore, that to say of anything whatever that we can mention that it absolutely is *not*, were to contradict ourselves: as if everything we can mention must be, must have some kind of being" (1953, p. 289).

Frege[4] de nominalizar todos os enunciados de nossa linguagem e adotar como único predicado a existência, ambos os problemas colapsam em um só. O enunciado "Temer não é jovem" se torna "A juventude de Temer não existe", um enunciado existencial negativo.

O problema dos nomes próprios vazios se revela particularmente agudo na medida em que o próprio interesse originário da filosofia pela categoria linguística de "nome" vem acompanhado de um interesse por resolver o paradoxo do discurso falso. De fato, mostrarei, em um primeiro momento, como Aristóteles introduz uma distinção entre "nome" e "declaração" justamente tendo em vista separar as condições de sentido das condições de verdade de um enunciado, abrindo assim a possibilidade para que um discurso seja falso sem ser, por isso, destituído de sentido. Em seguida, mostrarei como a ideia de que o sentido é anterior à verdade é radicalizada no *Tractatus* de Wittgenstein. Para o *Tractatus*, o sentido de uma proposição não apenas é anterior à sua verdade ou falsidade, mas é anterior à verdade ou falsidade de cada outra proposição. Uma das consequências desse movimento é a necessidade de se distinguir entre os nomes próprios assim chamados pela linguagem ordinária (como, p. ex., "Sócrates", "Júlio César"), que serão tratados como equivalentes a descrições, e nomes próprios genuínos, que nomeiam *objetos*, i.e., itens necessariamente existentes que compõem a substância do mundo. Em um terceiro momento, voltarei a atenção à obra de Kripke a fim de mostrar como este, ao negar que os nomes próprios da linguagem ordinária sejam equivalentes a descrições, vai chegar, em sua análise dos enunciados existenciais negativos, à recusa daquilo que Aristóteles e Wittgenstein punham como pressuposto, a saber, que o sentido de um enunciado seja anterior à sua verdade ou falsidade. Extrairei, a partir disso, algumas conclusões um tanto quanto paradoxais da análise de Kripke para estes enunciados. Por fim, tentarei fornecer uma via para entendê-las por meio de uma comparação, ainda que bastante breve, desses enunciados com o que Wittgenstein chama, na sua última obra, de proposições fulcrais (*hinge propositions*). A ideia central será a de que a "verdade" de sentenças como "Sherlock Holmes não existe" não corresponde a um *co-*

[4] Begriffschrift, §3.

nhecimento, mas a uma *segurança* ou *certeza*. Servimo-nos normalmente de um nome não porque *sabemos* que seu referente existe (ou não existe), mas porque estamos *seguros* a este respeito.

2. Sentido e Verdade em Aristóteles: a distinção entre "nome" e "declaração"

O tratado aristotélico *Da Interpretação* inicia-se do seguinte modo: "Inicialmente, é preciso estabelecer o que é nome (ὄνομα) e o que é verbo (ῥῆμα); em seguida, o que é negação (ἀπόφασις) e afirmação (κατάφασις), bem como declaração (ἀπόφανσις) e frase (λόγος)" (16a1-2). Para o presente propósito, será suficiente determo-nos no primeiro e no terceiro par de conceitos, a fim de traçarmos uma nova distinção, desta vez entre as noções de nome e declaração.

O nome, define Aristóteles, é "elocução com significado (σημαντικὴ) por convenção, sem tempo, da qual nenhuma parte é separadamente dotada de significado" (16a19-20). O verbo, em contraste, co-significa tempo. Mas o contraste para por aqui, pois o verbo partilha com o nome dessa característica de "simplicidade semântica", i.e., em ambos os casos, suas partes nada significam. Na verdade, o verbo, enunciado sozinho e em si mesmo – vai dizer Aristóteles – é também um nome e, quando usado em uma declaração, é um nome que co-significa um tempo particular. Em "Sócrates corre", por exemplo, "corre" é um verbo que significa "correr *agora*", mas o termo "correr", tomado isoladamente, é um nome do mesmo modo que "Sócrates" (é o nome de uma ação).

Já a frase "é enunciado com significação, da qual uma parte qualquer, separadamente, tem significado como enunciação, mas não como afirmação" (16b26-7). E dentre as frases, esclarece o estagirita, "nem toda frase é declarativa, mas apenas aquela em que ocorre pretender dizer algo verdadeiro ou falso; e isso não ocorre em qualquer frase; por exemplo, a prece é certamente uma frase, mas não é nem verdadeira nem falsa". Por fim, Aristóteles informa que é a frase declarativa (λόγος ἀποφαντικὸς) que o interessará mais de perto naquela obra.

Ora, o que se observa, na frase declarativa, na *proposição* (para usar um termo contemporâneo), é que, em contraste com o nome, aquilo que lhe é característico é que uma parte dela tem significado, mas não como afirmação (ou negação), ou seja, há uma parte dela que é um nome ou um verbo. Isto é, sua significação envolve uma "complexidade semântica"; ao contrário do nome, suas partes são separadamente dotadas de significado.

Aristóteles distingue, assim, dois sentidos do que é, para um item linguístico, "significar". Para um nome, significar é simbolizar algo, manter uma relação simples de nomeação com esse algo. Já para uma frase declarativa, significar é articular, é reunir de um certo modo coisas que simbolizam algo. Aristóteles dirá que há dois modos de se fazer essa reunião – a afirmação e a negação – e, a partir disso, distinguirá dois casos em que a proposição pode ser verdadeira e dois casos em que ela pode ser falsa, a depender tanto do modo pelo qual os nomes são reunidos na proposição quanto do modo pelo qual os nomeados se relacionam na realidade (se formam um composto ou se encontram-se separados).

Mas qual seria a motivação para a distinção entre dois tipos de significação? A resposta é clara: responder ao paradoxo do discurso falso. Pois se o discurso, para ser discurso, devesse apenas manter com as coisas uma relação simples, não haveria discurso falso, uma vez que a própria relação de significação já implicaria a verdade do discurso, ou melhor, a questão da verdade ou da falsidade do discurso sequer se colocaria. Sendo o falso e o verdadeiro uma questão de composição e separação (cf. 16a12), é necessário que o discurso seja essencialmente complexo para que a conclusão do paradoxo não se siga.

O que é tomado, então, como requerimento para a solução do paradoxo é o seguinte princípio: o sentido de uma frase declarativa (i.e., aquilo que a distingue de uma mera elocução, de um *flatus vocis*) deve ser anterior à sua verdade ou falsidade. Anterior não em sentido temporal, mas em sentido *quasi-epistêmico*: ainda que não saibamos se a declaração é verdadeira ou falsa, conhecemos o seu sentido, sabemos o que ela "quer dizer" (*quasi* pois quem sabe apenas o sentido de uma proposição, a rigor ainda não sabe absolutamente nada). E quando reconhecemos sua verdade ou falsidade, reconhece-

mos que há algo que permaneceu com isso inalterado, a saber, o seu sentido. O sentido da proposição é, deste modo, algo *rígido*, isto é, ele não vacila segundo o que é ou não o caso.

3. A radicalização do princípio da anterioridade do sentido no *Tractatus*

Não é algo controverso dizer que os elementos da solução Aristotélica para o paradoxo do falso aparecem com todas as suas letras no *Tractatus* de Wittgenstein. Em primeiro lugar, verdade ou falsidade é também lá uma questão de composição ou separação, de concatenação ou não-concatenação. Em segundo lugar, encontra-se também lá uma distinção forte entre dois sentidos do que é, para um item linguístico, significar: para o nome, significar é possuir uma *Bedeutung*, um referente; para a proposição, significar é figurar, articular nomes, exibir um sentido. Mas *algo* lá é diferente, pois, para começar, nunca seria possível fornecer, como exemplo de nome, "Sócrates", quanto menos "correr". Entender essa diferença em relação à teoria aristotélica requer entender o modo radical com que o princípio da anterioridade do sentido em relação à verdade é tomado na obra.

Para o *Tractatus*, não apenas o sentido de uma proposição é anterior a sua verdade ou falsidade, mas o sentido de cada proposição é anterior à verdade ou falsidade de cada outra proposição. O sentido de uma proposição não é, portanto, algo *rígido* apenas com respeito a sua verdade ou falsidade, mas ele é *rígido* seja lá o que aconteça no mundo. O espaço lógico todo é, desta forma, rígido. A verdade ou falsidade de cada proposição não altera em nada as possibilidades do que pode ou não vir a acontecer. Esse modo radical de entender a anterioridade do sentido em relação à verdade vai resultar no requisito de que as partes da proposição, isto é, os nomes, devam significar seja lá quais sejam os fatos do mundo e que, por conseguinte, o nome jamais pode nomear algo complexo, composto de partes. O requisito de que um nome não possua partes que signifiquem, proveniente da própria definição aristotélica de nome, transfigura-se agora em um requisito de que a coisa nomeada não contenha partes. Ao significado do nome, significado este essencialmente simples, sem partes, indivisível, Wittgenstein dá o nome de *objeto*. Nas *Observações*

Filosóficas, ele esclarece: "Aquilo que, em seu tempo, chamei de 'objeto', o simples, é simplesmente aquilo que posso designar sem ter de temer que ele talvez não exista, i. é, aquilo para o qual não há existência ou inexistência, e isso quer dizer: aquilo de que podemos falar *seja lá o que for o caso*" (PB: 36a).

E o que obriga Wittgenstein a realizar esse movimento? Isso parece-me ser o resultado de outros dois princípios:

> P_1. A lógica é *a priori* e, por isso, anterior a qualquer fato, a qualquer experiência possível.
> P_2. A lógica existe apenas *na* linguagem (do mesmo modo que, p. ex., a cor existe apenas em uma certa forma espacial), e por isso ela pressupõe que nomes tenham significado e proposições elementares tenham sentido (cf. 6.124).

Esses dois princípios aparecem notavelmente no seguinte trecho:

> A lógica é *anterior* a toda a experiência – de que algo é *assim*. Ela é anterior ao como, não é anterior ao quê. / E se não fosse assim, como poderíamos aplicar a lógica? Poder-se-ia dizer: se houvesse uma lógica ainda que não houvesse um mundo, como poderia então haver uma lógica, já que há um mundo? (*Tractatus*, 5.552-5.5521)

Como resultado, o princípio da anterioridade do sentido em relação à verdade, no *Tractatus*, conduz a uma hierarquia estrita entre três níveis. Em primeiro lugar, há o nível que condiciona o entendimento da lógica, a saber, o nível do "Quê" (Was), que corresponde, na linguagem, ao nível dos nomes e proposições elementares. Em seguida, há o nível da lógica, dependente do primeiro, mas ainda inteiramente *a priori*. Por fim, há o nível do *a posteriori*, do "Como" (*Wie*), que corresponde, na linguagem, às proposições verdadeiras. Três domínios, portanto: o que condiciona o entendimento do *a priori*, o *a priori* e o *a posteriori*; em outros termos: metafísica, lógica e ciência. A hierarquia diz respeito ao fato de que nenhum dos níveis posteriores condiciona ou afeta os anteriores. Uma descoberta científica jamais poderia produzir efeitos na lógica, muito menos na metafísica; uma "descoberta" na lógica jamais po-

deria produzir efeitos naquilo que condiciona o entendimento da lógica, isto é, na substância do mundo.

Na superfície de todo este arcabouço conceitual se encontra a *nossa* linguagem, a linguagem em que nomes ordinários não denotam objetos necessariamente existentes, que atribui o mesmo sinal a símbolos diferentes, enfim, que comete todo tipo de atrocidade ao olhar da lógica. Desta linguagem, o *Tractatus* não tem muito a dizer, a não ser i) que ela está logicamente correta e que ii) ela é um traje que disfarça o pensamento. Está logicamente correta, na verdade, por ter algo por baixo do traje, algo que, em sua forma pura, seria o pensamento nu e cru. Os nomes próprios da linguagem ordinária fazem parte do traje; os nomes genuínos, do pensamento. Consequentemente, os nomes próprios ordinários não fazem parte da análise final da sentença, sendo supostamente analisados e desaparecendo ao fim da análise.

4. Kripke e o problema dos enunciados existenciais negativos

Em *Naming and Necessity*, Kripke critica as teorias descritivistas dos nomes ordinários, mostrando que estes não são semanticamente equivalentes a descrições. No artigo *Vacuous Names and Fictional Entities* (1973a/2011) e de modo mais detalhado nas *John Locke Lectures* de 1973, posteriormente publicadas com o título *Reference and Existence* (1973b/2013), ele amplia suas considerações para o caso dos nomes vazios (particularmente no caso de nomes que ocorrem na ficção), abordando a questão dos enunciados existenciais negativos.

Em linhas gerais, o argumento de Kripke contra a solução descritivista para o caso dos nomes vazios é o seguinte: se a teoria descritivista fosse correta, isso resultaria em que as sentenças que contêm o nome "Sherlock Holmes" enunciariam algo a respeito de uma entidade que seria, digamos, "holmesca" (isto é, que teria todas ou ao menos boa parte das qualidades descritas por Arthur Conan Doyle sobre Sherlock Holmes), e que seria a única entidade "holmesca". Todavia, ainda que haja unicamente alguém "holmesco", isso não significa que a ficção seja sobre essa pessoa. Kripke menciona o exemplo de autores que alertam que os nomes usados na obra são ficcionais, e que qual-

quer semelhança com pessoas reais é mera coincidência (1973b/2013, p. 26-7); caso a teoria descritivista fosse correta, esse aviso ao leitor seria claramente um disparate. Além disso, pode ser o caso de a ficção escrita por Conan Doyle ser realmente sobre alguém existente e esta pessoa não ser "holmesca". Basta supor que o autor estivesse escrevendo uma história ficcional sobre uma pessoa real (como ocorre no caso de Napoleão em *Guerra e Paz* de Tolstói) e que, no entanto, tenha inventado uma série de características que são, em realidade, falsas quando aplicadas à Napoleão.

Agora, dado que Conan Doyle *não* estava escrevendo sobre uma pessoa real quando escrevia as histórias de Holmes, será que poderíamos dizer que Holmes, embora não exista, poderia existir? A resposta, para Kripke, é negativa. Kripke argumenta que não haveria critérios para decidir em quais situações ele teria existido. Pois se o nome não equivale a uma descrição, de qual critério disporíamos para afirmar que alguém *seria* Holmes em tal e tal situação? Nesse sentido, o nome "Holmes" é similar a "Bandersnatch": dado que Lewis Carroll não estava se referindo a nenhuma espécie existente quando usou o nome "Bandersnatch", não há nenhum critério para nos referirmos a uma situação em que bandersnatches existem, e o mesmo ocorre com "Holmes".

Observem, por outro lado, que o fato de Conan Doyle ter ou não escrito sobre uma pessoa real é algo empírico: pode ocorrer de descobrirmos que Conan Doyle de fato escreveu sobre (e não meramente se inspirou em) uma pessoa real. Se um baú nos é revelado com diários secretos de Conan Doyle, os quais continham informações sobre quem seria esta pessoa a que chamamos de Sherlock Holmes, o que ela fazia e como as suas histórias se vinculam à sua vida etc., essa descoberta, certamente empírica, forneceria evidência suficiente para tomarmos por certo aquilo que antes considerávamos inimaginável, a saber, a existência de Sherlock Holmes. O que é indigesto, em um primeiro momento, é que este algo empírico, de certa forma, parece alterar o espaço de possibilidades do que falamos sobre Sherlock Holmes. Pois, se Conan Doyle tivesse em mente uma pessoa real quando escreveu sobre sua personagem, a existência dessa personagem seria um fato contingente e, neste

caso, Sherlock Holmes claramente poderia não ter existido. Tudo se passa então como se ocorresse, aqui, uma espécie de assimetria modal-existencial: se Sherlock Holmes existe, ele poderia não existir, mas se ele não existe, ele não poderia existir. A descoberta acima referida parece ser, portanto, o caso de uma descoberta empírica que alteraria a lógica. Isso ficará mais claro logo em diante, quando expormos o que Kripke diz sobre os enunciados existenciais negativos. Antes disso, porém, será útil expor de maneira breve o que ele diz sobre nomes de personagens fictícios em geral.

Segundo Kripke, devemos distinguir três usos distintos de nomes de personagens fictícios:

(1) Uso do autor durante a escrita da obra ou do ator durante a apresentação da peça. Nestes casos, não se faz um uso de nomes próprios para se referir a algo real; ele apenas *faz de conta* (*pretend*) que usa o nome próprio para referir; da mesma maneira, na obra fictícia, não há realmente proposições, o autor apenas *faz de conta* que expressa proposições. Kripke chama este princípio sobre as proposições e nomes na ficção de "Princípio do Faz-de-Conta" (R&E, p. 24).

(2) Uso *meta-ficcional*, que ocorre quando o nome é usado para descrever fatos *reais* sobre a ficção. Ex.: "Sherlock Holmes foi criado por Arthur Conan Doyle". Neste caso, diz Kripke, o nome fictício se refere a um objeto abstrato, objeto cuja existência, porém, depende de atividades empíricas (*e.g.*, escrever obras de literatura, peças de teatro, etc.).

(3) Uso *objeto-ficcional*[5], que ocorre quando alguém usa o nome de um personagem fictício para descrever o que é verdadeiro no interior da ficção. Ex.: "Sherlock Holmes era fumante". Nestes casos, o nome fictício continua a se referir ao mesmo objeto abstrato do caso anterior, mas o predicado é entendido de modo *não-standard* (*e.g.*, fumante-na-ficção, detetive-na-ficção, etc.).

Agora podemos retornar ao problema dos enunciados existenciais negativos (no nosso exemplo, ao enunciado "Sherlock Holmes não existe"). Na

[5] Os termos "meta-ficcional" e "objeto-ficcional" foram introduzidos por Nathan Salmon (1998).

última aula do curso de 1973, Kripke inicia sua abordagem do problema pela recusa de três soluções. Na primeira delas, o nome "Sherlock Holmes" designaria um personagem fictício (um objeto abstrato) e, com isto, resolveríamos o problema do "objeto" do discurso. A solução é recusada pois, neste caso, a proposição existencial negativa seria falsa, pois Sherlock Holmes, enquanto um personagem fictício, existe. Na segunda delas, o enunciado é interpretado de maneira a significar a proposição "Sherlock Holmes não é alguém real, mas apenas um personagem fictício", mais ou menos à maneira como dizemos, do pato da Fiesp, que ele não é um pato de verdade. Kripke ataca essa solução, sendo seu principal argumento o de que uma sentença como "Suponha que Napoleão não existe" não quer dizer "Suponha que Napoleão não seja alguém real, mas meramente um personagem fictício". Uma pessoa real – argumenta Kripke – pode não existir, mas não pode ser um personagem fictício, assim como não pode ser, digamos, um número primo. A última solução recusada por Kripke é a análise metalinguística. Segundo essa análise, o enunciado problemático seria interpretado de maneira a significar a proposição "O nome 'Sherlock Holmes' é um nome vazio". A crítica de Kripke a essa solução é que ela não funciona em situações contrafactuais, uma vez que, p. ex., Napoleão poderia muito bem ter existido ainda que o nome "Napoleão" fosse um nome vazio. Ou seja, quando descrevemos as circunstâncias em que o nome "Napoleão" seria vazio, isto não é o mesmo que descrever as circunstâncias em que Napoleão não teria existido.

Qual é, então, a solução de Kripke para o problema dos enunciados existenciais negativos? Antes de se pronunciar sobre ela, Kripke avisa:

> Parece que, em algum sentido, a análise de um enunciado singular de existência dependerá de se o enunciado é verdadeiro. E isto, é claro, parece ser em si e por si mesmo absolutamente intolerável: a análise de um enunciado não deveria depender de seu valor de verdade. Ou, em todo caso, esse pode ser nosso preconceito. Contudo, não se deveria, aqui, ignorar os fatos, e os fatos parecem ser, para mim, tais como eu os afirmei. (KRIPKE 1973b/2013, p. 147)

De fato, a análise proposta por Kripke de enunciados singulares existenciais não preserva o princípio da anterioridade do sentido em relação à verdade. Tomemos, p. ex., o enunciado "Sherlock Holmes existe". Se este

enunciado é verdadeiro, sua análise é simplesmente E(Sherlock Holmes), isto é, a atribuição de um predicado (a existência) a algo existente. Vale lembrar que, apesar de ser verdadeiro, o enunciado não é necessariamente verdadeiro, uma vez que a existência de Sherlock Holmes é contingente. Por outro lado, se o enunciado é falso, ele não pode ser concebido de modo a predicar uma propriedade de algo que não existe (pois "algo que não existe" é uma contradição em termos). Nesse caso, Kripke propõe analisar o enunciado verdadeiro correspondente, a saber, "Sherlock Holmes não existe" como "Não há uma proposição (verdadeira) que afirme que Sherlock Holmes existe".

Um corolário imediato da solução de Kripke é que, quando estamos diante de uma questão sobre um enunciado existencial (e.g.: "Será que Vulcano existe?"), não há um sentido (um *Gedanke*) associado a esta questão, mas responder afirmativamente à questão (ou então crer em sua verdade) significa algo distinto de responder negativamente à questão (ou então crer em sua falsidade). Essa consequência é expressa de maneira cristalina por Kripke quando este diz que:

> Aquele que disse que Moisés existe acreditava estar pronunciando uma proposição verdadeira e eu penso que ele de fato estava. Aquele que negou isso, porém, na minha visão, não estava negando aquela proposição – isso se ele for um filósofo correto, o que provavelmente não é o caso. Ao invés disso, ele pensava que não havia proposição verdadeira para ser enunciada. (KRIPKE 1973b/2013, p. 159)

Aquele que acredita, portanto, que Sherlock Holmes não existe, acredita, segundo a análise proposta por Kripke, que não há uma proposição que afirme que Sherlock Holmes existe. Isto é, a crença se revela, em um certo sentido, como uma crença *sobre* a linguagem (mais precisamente, sobre quais proposições na linguagem existem ou não existem). Do mesmo modo, aquele que acreditou que Vulcano (o planeta hipotético) existia mantinha uma crença sobre a linguagem, a saber, que um nome próprio havia sido propriamente introduzido. Quando se chegou à conclusão de que Vulcano não existia, o que foi descoberto, na verdade, não é que um objeto possuía a estranha propriedade de não existir, mas que as condições necessárias para a introdução de um nome na linguagem, condições sem as quais não pode haver nomeação algu-

ma, não eram satisfeitas. Não seria um equívoco chamar tal descoberta de *linguística*: descobre-se que um pseudo-nome não é, na verdade, um nome.

Isso indica que, na análise de Kripke, há ainda espaço para uma distinção, presente no *Tractatus*, entre um nome próprio genuíno e um nome próprio aparente. A diferença é que o critério para essa distinção não é mais *metafísico*: não é por denotar um objeto simples, mas sim por denotar um objeto existente que um termo é genuinamente um nome. Por outro lado, isso não significa que a proposição que afirma a existência do objeto nomeado é condição de sentido ou de verdade das proposições que usam o nome pois, como exposto acima, a proposição que afirma a existência tem seu sentido determinado somente quando se sabe seu valor de verdade e, *precisamente por essa razão*, ela nunca poderia funcionar como condicionante do que quer que seja. Kripke pode, então, recusar a conclusão do *Tractatus* de que um nome não pode se referir a algo que poderia não existir precisamente porque a atribuição da existência, para ele, não significa que as partes de um complexo estão juntas (de modo que poderiam estar juntas caso não estivessem, e poderiam não estar juntas caso estivessem, obedecendo deste modo uma simetria modal-existencial). O que ocorre, simplesmente, é que a existência do nomeado não é propriamente descrita por uma proposição que possui sentido anteriormente a sua verdade ou falsidade.

Na Seção seguinte, pretendemos tentar jogar luz nestas conclusões um tanto quanto paradoxais às quais a análise de Kripke nos conduz por meio de uma comparação, ainda que breve, destes enunciados existenciais singulares com aquilo que Wittgenstein chama, em sua última obra, de proposições fulcrais (*hinge propositions)*.

5. Proposições existenciais e proposições fulcrais

Parece-me iluminador traçar uma comparação dos enunciados que estão no centro de nossos problemas com uma classe de proposições que Wittgenstein, em *Sobre a Certeza*, chama de fulcrais.

Em primeiro lugar, assim como a solução de Kripke para o problema dos enunciados existenciais negativos sugere que os enunciados singulares de existência não obedecem ao princípio da anterioridade do sentido em relação à verdade, deve-se notar, também nas proposições fulcrais, uma certa dependência estrita entre sentido e verdade. A este respeito, Wittgenstein diz: "Testa-se na *verdade* das minhas declarações a minha *compreensão* delas. / Ou seja: se fizer certas declarações falsas, torna-se duvidoso se eu as compreendo" (UG: 80-1). Se alguém diz, por exemplo, que houve um homem que já foi ao Sol, torna-se duvidoso se ele compreende o significado de "Sol"; do mesmo modo, se alguém diz que tentou se encontrar com Sherlock Holmes para pedir-lhe um autógrafo, torna-se duvidoso se ele compreendeu o significado de "Sherlock Holmes".

Em segundo lugar, encontra-se, em *Sobre a Certeza*, a ideia de que as proposições fulcrais não fundamentam as nossas outras crenças por serem proposições *conhecidas*. Não se trata, aqui, de conhecimento, de um "ver que é verdadeiro". Trata-se, ao invés disso – dirá Wittgenstein – de um agir. A meu ver, temos a mesmíssima situação no caso da sentença "Sherlock Holmes não existe". Lembremos que, na análise de Kripke, essa sentença verdadeira diz: "Não há proposição que afirme que Sherlock Holmes existe". Aqui fica claro que não há nada para ser conhecido sobre o mundo. A existência de Sherlock Holmes não é uma possibilidade que é descartada uma vez que sabemos a verdade da sentença acima. Pelo contrário: se compreendemos a sentença acima, também compreendemos que Sherlock Holmes nunca poderia ter existido, ou melhor, que a sentença "Sherlock Holmes existe" não expressa uma possibilidade real. Deste modo, nossa compreensão comum das sentenças sobre Sherlock Holmes não depende de uma crença ou de um conhecimento comum que todos nós compartilhamos. Ao invés disso, essa compreensão comum se manifesta em um *agir comum*. Por exemplo, se al-

guém diz "Sherlock Holmes era fumante" e uma outra pessoa pega um livro em que Conan Doyle atesta isso, isso é suficiente para ambas concordarem, e se alguém diz que conquistou o coração de Sherlock Holmes, ou que o encontrou ali na esquina, isso é suficiente para que procuremos dar um sentido *não-standard* ao que ela disse (e isso se manifestaria no tipo de *atitude* com a qual reagiríamos a uma sentença deste tipo).

Em terceiro lugar, o caso em um baú nos é revelado com diários secretos de Conan Doyle parece poder ser comparado ao caso em que as fronteiras entre as proposições fulcrais e as proposições "em fluxo" (para usar a metáfora do rio) começam a se tornar difusas. Pois aqui surge precisamente um contexto no qual uma dúvida a respeito da existência ou não existência de Sherlock Holmes passa a ser "concebível". Mas vale insistir que este caso não é o caso em que uma questão, posta pela linguagem, estaria recebendo uma resposta positiva ou negativa. Trata-se aqui, justamente, de decidir se as condições para a nomeação estavam ou não presentes. Trata-se aqui de decidir se a preparação da linguagem foi adequada (se a rainha branca, p. ex., foi colocada na casa branca). E aqui o critério de adequação é mais ou menos arbitrário (assim como também há o arbitrário na ciência): não poderia o baú ser uma farsa? Quem decide sobre o que é firme e o que é contestável? (cf. UG: 125)

Se nos lembrarmos que a distinção fundamental que está sendo traçada em *Sobre a Certeza* é entre as noções de conhecimento (*Wissen*) e de segurança (*Sicherheit*) ou de certeza (*Gewißheit*), o que se propõe aqui é que o uso de um nome seja vinculado não a um *conhecimento* da existência do nomeado, o que tornaria o sentido das proposições que usam o nome dependentes de um pressuposto de existência, mas a uma segurança, a uma certeza. O próprio Wittgenstein parece sugerir que tomemos esta via. Após contrastar ambas as noções dizendo que, quando me sirvo imediatamente de algo (de uma toalha, p. ex.), esse servir-se imediatamente corresponde a uma *segurança*, não a um saber, Wittgenstein coloca a seguinte questão (retórica, a meu ver): "Mas não me sirvo do nome de uma coisa assim também"? (UG: 511).

Referências

ANGIONI, Lucas. *Introdução à Teoria da Predicação em Aristóteles*. Campinas/SP: Editora Unicamp, 2006.

FREGE, Gottlob. *Begriffsschrift und andere Aufsätze*, Hildesheim, Zürich, New York: Georg Olms, 2007. (abreviado como *Begriffsschrift)*

KRIPKE, Saul, *Naming and Necessity*, Cambridge, MA: Harvard University Press, 1980.

_____. *Reference and Existence*: The John Locke Lectures. Oxford, New York: Oxford University Press, 2013. (abreviado como R&E, citado como 1973b/2013)

_____. "Vacuous Names and Fictional Entities". In S. Kripke, *Philosophical Troubles*: Collected Papers, Volume 1. Oxford University Press, 2011, p. 52–74. (citado como 1973a/2013)

MOORE, George. *Some Main Problems of Philosophy*, London: George Allen & Unwin, 1953.

RUSSELL, Bertrand. "On Denoting". *Mind*, 14.56, 1906, p. 479–493.

SALMON, Nathan. "Nonexistence". *Noûs*. 32.3, 1998, p. 277-319.

WITTGENSTEIN, Ludwig. *Philosophische Bemerkungen*, ed. por Rush Rhees, Frankfurt: Suhrkamp, 1964. (abreviado como PB, tradução minha)

_____. *Tractatus Logico-Philosophicus*, Tradução: Luis Henrique Lopes dos Santos. São Paulo: EdUSP, 1993. (abreviado como *Tractatus*)

_____. *Da certeza*. Tradução: Maria Elisa Costa. Lisboa: Edições 70, 2012.

TESE DE SUSZKO E PLURALISMO LÓGICO

Edson Bezerra[1]

RESUMO

A Tese de Suszko (SUSZKO, 1977) afirma a existência de somente dois valores lógicos, o verdadeiro lógico, o verdadeiro e o falso. Neste artigo, argumento em favor da compatibilidade da Tese de Suszko com a tese do Pluralismo Lógico tal como formulado por Hjortland (HJORTLAND, 2011).

Palavras-chave: Lógicas Multivaloradas. Tese de Suszko. Pluralismo Lógico. Consequência Lógica.

ABSTRACT

Suszko's Thesis asserts the existence of only two logical values, the true and the false. In this paper, I argue in favor of the compatibility of Suszko's Thesis and the version of Logical Pluralism formulated by Hjortland (HJORTLAND, 2011).

Keywords: Many-valued logics. Suszko's Thesis. Logical Pluralism. Logical Consequence.

1. Introdução

Neste artigo, argumentarei em favor da compatibilidade da Tese de Suszko com a tese do pluralismo lógico, que defende a existência de mais de uma lógica correta. As lógicas multivaloradas são lógicas que diferem da lógica clássica por admitirem valores além dos valores verdadeiro e falso. De acordo com Kneale & Kneale (KNEALE & KNEALE, 1962), tais lógicas surgiram a partir dos trabalhos independentes de Łukasiewicz, Post e Bernays. Bernays, em 1919, utiliza tabelas de verdade multivaloradas para provar a independência dos axiomas da lógica proposicional clássica. Embora seja discutível se Bernays pode ser considerado como um dos pioneiros das lógicas multivaloradas, uma vez que ele usou tabelas multivaloradas somente como

[1] Mestre e doutorando da Universidade Estadual de Campinas (Unicamp). Este trabalho foi financiado pela Coordenação de Aperfeiçoamento de Nível Superior (CAPES), código 001. E-mail: edson.vinber92@gmail.com. ORCID: 0000-0002-0865-5197.

técnicas para provar um resultado metateórico, sua contribuição para as investigações posteriores em lógicas multivaloradas é indiscutível.[2] As investigações de Łukasiewicz, em 1920, acerca do problema dos futuros contingentes deram origem a uma série de lógicas multivaloradas. Segundo ele, enunciados sobre o futuro não podem ser verdadeiros nem falsos, mas *possíveis.* Uma das lógicas resultantes dessa análise é a lógica trivalorada $Ł_3$, que inclui o *possível* como um valor de verdade. Em 1921, Post introduz o método de tabelas de verdade como um procedimento de decisão para lógicas proposicionais. Dada a generalidade de sua investigação, ela abrange casos arbitrários de *n* valores. De maneira geral, podemos dizer que as lógicas multivaloradas podem ser vistas não somente como lógicas, mas também como um método de investigação metateórica. Uma vantagem dessas lógicas é a de preservarem o aspecto composicional em sua semântica, o que facilita a demonstração de resultados de correção e completude.

A *Lógica do Paradoxo* (LP) (PRIEST, 1979) é uma lógica multivalorada frequentemente utilizada nas investigações dos paradoxos semânticos, dentre os quais o *paradoxo do mentiroso* é um dos mais conhecidos. Tarski (TARSKI, 1977) mostra que a introdução de um predicado *Tr*() de sentença verdadeira na linguagem objeto de uma teoria **T** capaz de expressar aritmética permite a formulação da sentença do mentiroso λ, que assere sua própria falsidade. Ademais, a presença de tal sentença acarreta uma contradição da forma *Tr*([λ]) *e não-Tr*([λ]). Se **T** for uma teoria que possui a lógica clássica como sua base, então **T** é trivial em virtude do *Princípio da Explosão* (*PE,* $\varphi, \neg\varphi \vDash_{\mathbf{T}} \psi$). Por outro lado, se **T** tiver LP como lógica de base, então a contradição acarretada por λ não implica trivialidade.[3] Por conseguir lidar não trivialmente com contradições, LP é uma boa candidata para ser base de teorias que possuem contradições, mas que não são triviais.

Embora LP fora escolhida como exemplo, existe uma variedade de casos onde lógicas multivaloradas são utilizadas como ferramentas de análise de diversos problemas, tanto de natureza puramente formal quanto os de natu-

[2] Como Carnielli (CARNIELLI, 2012) ressalta, os trabalhos de Bernays contribuíram fortemente para o desenvolvimento das semânticas algébricas.

[3] Em (PRIEST, 2006a), o leitor encontrará propostas de aplicações de LP em teorias contraditórias, tais como a própria teoria cantoriana de conjuntos.

reza filosófica. O sucesso dessas lógicas pode sugerir a existência de mais valores de verdade além de verdade e falsidade. Nesse contexto, a *Tese de Suszko* surge como uma resposta a essa questão. Partindo da constatação de que o conjunto de valores de verdade dessas lógicas é bipartido, para que seja possível definir os conceitos de tautologia e consequência lógica, Suszko (SUSZKO, 1977) afirma que existem somente dois *valores lógicos*: *verdadeiro* e *falso*. Os muitos valores de verdade dessas lógicas são entendidos por ele como *valores algébricos*, entendidos como sendo os referentes admissíveis das fórmulas. Como contraparte técnica à tese filosófica, Suszko mostra ser possível apresentar uma semântica de bivalorações para essas lógicas, mostrando o caráter bivalente dessas lógicas de um ponto de vista técnico.

A possibilidade de reduzir os muitos valores algébricos aos valores lógicos verdade e falsidade pode sugerir que as lógicas multivaloradas não sejam interessantes de um ponto de vista filosófico e que, além disso, o caráter bivalente dessas lógicas implica a existência de somente uma lógica, a lógica clássica. Neste artigo, argumentarei que isso não é o caso. Ou seja, argumentarei que o caráter bipartite dessas lógicas não implica no monismo lógico. Em particular, argumentarei em defesa da compatibilidade da Tese de Suszko com a proposta do pluralismo lógico tal como formulado por Hjortland (HJORTLAND, 2011).

2. Lógicas Multivaloradas e Tese de Suszko.

Nesta seção, apresentarei a teoria das matrizes lógicas, tendo a lógica trivalorada LP (PRIEST, 1979) como objeto de análise. Em seguida, apresentarei a Tese de Suszko e um resultado de redução a bivalorações delineado por Suszko em (SUSZKO, 1977) para o caso de LP. Apresentarei uma possível objeção à Tese de Suszko bem como uma possível resposta. Por fim, argumentarei em defesa da plausibilidade da Tese de Suszko.

2.1. Lógicas Multivaloradas: linguagem e semântica

Nesta seção introduziremos informalmente alguns conceitos básicos de lógica proposicionais. Seja **L** uma lógica. A linguagem L de **L** é um con-

junto $L = \{At, \neg, \rightarrow\}$ onde $At = \{p_i \mid i \in \mathbb{N}\}$ é um conjunto de variáveis proposicionais, ou proposições atômicas, $\neg$ é o conectivo unário de negação e $\rightarrow$ é o conectivo binário de implicação.[4] O conjunto de fórmulas For^L é definido como segue. Para $\varphi, \psi \in For^L$:[5]

$$p_i \mid \neg\varphi \mid \varphi \rightarrow \psi$$

A semântica para as lógicas multivaloradas é definida a partir de uma generalização da semântica de tabelas de verdade para a lógica proposicional clássica. Essa generalização deu origem à teoria das *matrizes lógicas*, que foi desenvolvida a partir dos trabalhos de Łukasiewicz, Tarski e Lindenbaum (TARSKI, 1977).[6] Uma interpretação para a linguagem L é uma matriz $\mathscr{M} = (V, \neg, \rightarrow, D)$, onde V é um conjunto de valores algébricos, $\neg, \rightarrow$ são operações sobre V, e $D \subset V$ o conjunto de valores *distinguidos.*[7] Uma valoração é uma função $v: At \rightarrow V$ do conjunto de variáveis proposicionais no conjunto V de valores algébricos. A valoração v é estendida recursivamente para $v: For^L \rightarrow V$ como segue. Sejam f e h duas funções, então: $v(\neg\varphi) = f(v(\varphi))$; $v(\varphi \rightarrow \psi) = h(v(\varphi), v(\psi))$. As definições de f e h são especificadas de acordo com a lógica. O conjunto de todas as valorações $v: For^L \rightarrow V$ constitui a semântica de L, denotado por $sem_{\mathbf{L}}$. Dizemos que uma valoração $v \in sem_{\mathbf{L}}$ é modelo para fórmula $\varphi \in For^L$ se $v(\varphi) \in D$. Seja $\Gamma \subseteq For^L$ um conjunto de fórmulas. Dizemos que v é modelo de Γ se $v(\varphi) \in D$, para todo $\varphi \in \Gamma$. Dizemos que φ é *consistente* se $v(\varphi) \in D$, para alguma $v \in sem_{\mathbf{L}}$. Dizemos que φ é uma *tautologia* se $v(\varphi) \in D$, para toda $v \in sem_{\mathbf{L}}$. Dizemos que φ é uma *contradição* se $v(\varphi) \notin D$, para toda $v \in sem_{\mathbf{L}}$. A relação $\vDash_{\mathbf{L}} \subseteq \wp(For^L) \times For^L$ de consequência lógica é

[4] Neste trabalho optamos por utilizar linguagens que possuem somente implicação e negação como primitivos, uma vez que esses conectivos tomados como primitivos abrangem uma classe considerável de lógicas multivaloradas. Claramente, uma generalização para quaisquer conectivos é imediata. Para uma apresentação mais geral, convidamos o leitor a consultar (WÓJCICKI, 2013).

[5] Daqui em diante utilizaremos as variáveis proposicionais sem seus índices por simplicidade de escrita.

[6] O artigo onde a teoria das matrizes lógicas é apresentada se chama *Investigations into the Sentential Calculus*, p. 38 – 60.

[7] Propositalmente, utilizamos os mesmos símbolos para os conectivos e as operações para ressaltar que as operações são as contrapartes semânticas dos conectivos.

definida como segue: dizemos que $\Gamma \vDash_{\mathbf{L}} \varphi$ se e somente se todo modelo de Γ é também modelo de φ. De acordo com Marcos (MARCOS, 2009), as matrizes lógicas preservam a *verofuncionalidade* das lógicas multivaloradas. Uma lógica é *verofuncional* se sua semântica é tal que o valor de verdade de uma fórmula complexa é determinado pelos valores de verdade das fórmulas simples que a compõem.

Dizemos que uma lógica **L** é multivalorada se a cardinalidade do conjunto V é maior que dois. Por outro lado, devemos prestar atenção ao seguinte detalhe. Pode ser o caso que uma mesma lógica **L** seja caracterizada por matrizes $\mathscr{M}_1 = (V_1, \neg, \rightarrow, D_1)$ e $\mathscr{M}_2 = (V_2, \neg, \rightarrow, D_2)$ tais que V_1 possui n elementos e V_2 possui m elementos tais que n e m são diferentes. Assim, a mesma lógica possuiria n e m valores, o que pode soar absurdo. Isso se deve ao fato de que a uma lógica não está associada somente uma matriz, mas uma classe de matrizes. Para resolver esse aparente dilema, Marcos (MARCOS, 2009) chama uma lógica **L** de *genuinamente n*-valorada se n é a menor cardinalidade do conjunto V da matriz $\mathscr{M} = (V, \neg, \rightarrow, D)$ associada à lógica **L**. Dada essas definições, podemos considerar uma lógica **L** como um par $\mathbf{L} = (L, \vDash_{\mathbf{L}})$ onde L é a linguagem de **L** e $\vDash_{\mathbf{L}}$ é a relação de consequência semântica definida no conjunto de fórmulas For^{L}.[8] Como podemos ver, a distinção entre os valores distinguidos e não-distinguidos é necessária para que os conceitos de tautologia e consequência lógica sejam definidos.

De acordo com Łoś & Suszko (ŁOŚ & SUSZKO, 1958), a definição da relação de consequência lógica como preservação de valor de distinguidos implica que a relação $\vDash_{\mathbf{L}}$ tenha as seguintes propriedades:

(*I*) $\Gamma \cup \{\varphi\} \vDash_{\mathbf{L}} \varphi$ (*Reflexividade*)

(*II*) Se $\Gamma \vDash_{\mathbf{L}} \varphi$ e $\Gamma \subseteq \Delta$, então $\Delta \vDash_{\mathbf{L}} \varphi$ (*Monotonicidade*)

(*III*) Se $\Gamma \vDash_{\mathbf{L}} \varphi$ e $\Delta, \varphi \vDash_{\mathbf{L}} \psi$, então $\Delta, \Gamma \vDash_{\mathbf{L}} \psi$ (*Corte*)

[8] Claramente, uma definição de lógica mais geral poderia ser dada, não se restringindo à consequência semântica. De fato, é mais usual trabalhar com uma relação de consequência geral $\Vdash$. Aqui, por questões de simplicidade, optamos por trabalhar somente com a relação de consequência semântica $\vDash$. De todo modo, ambos $\Vdash$ e $\vDash$ são sempre definidos a partir do conjunto de fórmulas de uma lógica **L**. Uma exposição mais geral da relação de consequência pode ser encontrada em (MALINOWSKI, 1993).

A demonstração desses enunciados é simples. A reflexividade garante que qualquer fórmula é consequência lógica de si própria. A monotonicidade preserva a consequência lógica mediante superconjuntos. E o corte garante, de certo modo, a transitividade da relação de consequência. Dizemos que uma relação de consequência é *tarskiana* se satisfaz as propriedades (*I*)-(*III*). Considere agora um *endomorfismo* ε: $At \rightarrow For^{L}$ que atribui para cada variável proposicional uma fórmula. Um endomorfismo é basicamente um mapeamento de uma estrutura em si própria. Claramente, ε pode ser estendido para todas as fórmulas, de modo que ε: $For^{L} \rightarrow For^{L}$. Seja Γ um conjunto de fórmulas. Por ε(Γ) entendemos o conjunto ε(Γ) = {ε(γ) | γ ∈ Γ}, obtido a partir da aplicação do endomorfismo ε nas fórmulas do conjunto Γ. Se além das propriedades (*I*) – (*III*), $\vDash_{\mathbf{L}}$ satisfizer a propriedade

(*IV*) Γ $\vDash_{\mathbf{L}}$ φ se e somente se (sse) ε(Γ) $\vDash_{\mathbf{L}}$ ε(φ) (*Estruturalidade*)

ela é chamada de *tarskiana estrutural*. Mostraremos agora que $\vDash_{L}$ é estrutural:

Teorema 2.1. $\vDash_{L}$ é estrutural.

Prova: Assuma Γ $\vDash_{\mathbf{L}}$ φ e suponha ε(Γ) $\nvDash_{L}$ ε(φ). Então existe uma valoração $v \in sem_{\mathbf{L}}$ tal que $v(\varepsilon(\gamma)) \in D$, para todo ε(γ) ∈ ε(Γ), e $v(\varepsilon(\varphi)) \notin D$. Uma vez que ε é um endomorfismo em *L*, $v(\gamma) \in D$, para todo γ ∈ Γ, e $v(\varphi) \notin D$. O que contradiz nossa assunção inicial. Logo, ε(Γ) $\vDash_{\mathbf{L}}$ ε(φ). O raciocínio converso é similar. *Q.E.D.*

A estruturalidade preserva a relação de consequência mediante substituição *uniforme* das variáveis proposicionais que ocorrem nas fórmulas de Γ e em φ. Ou seja, a definição *bipartite* do conjunto *V* de valores de verdade da matriz $\mathscr{M}$ determina que $\vDash_{\mathbf{L}}$ seja tarskiana estrutural. A definição da matriz é bipartite no sentido de todo valor de verdade ser ou não distinguido: para todo v ∈ *V*, v ∈ *D* ou v ∉ *D*. Considere o seguinte exemplo:

Exemplo 2.2. A *Lógica do Paradoxo* (LP) (PRIEST, 1979) é apresentada a partir da matriz $\mathscr{M} = (\{1,½,0\}, \neg, \rightarrow, \{1,½\})$ cujas operações possuem as seguintes tabelas de verdade:

	¬
1	0
½	½
0	1

→	1	½	0
1	1	½	0
½	1	½	½
0	1	1	1

As valorações de LP são definidas como segue:

$v(\neg\varphi) = 1 - v(\varphi)$
$v(\varphi \rightarrow \psi) = \text{máximo}(1 - v(\varphi), v(\psi))$

O conjunto de todas as valorações v: $For^{L} \rightarrow \{1,½,0\}$ constitui a semântica de LP. A definição da relação $\vDash_{LP}$ é similar à definição que demos acima. Não apresentaremos aqui um sistema dedutivo para LP por não ser necessário para os nossos propósitos. O leitor encontrará um sistema de tablôs completo e correto em relação à matriz $\mathscr{M}$ de LP em (PRIEST, 2008). Uma vez que a definição de $\vDash_{LP}$ como preservando valores distinguidos, as propriedades de $\vDash_{LP}$ são exatamente as descritas acima. Ou seja, $\vDash_{LP}$ é reflexiva, monotônica, possui corte e é estrutural. A enumeração dessas propriedades será importante ao abordamos adiante o resultado de redução de Suszko.

2.2. Dois valores: Tese e Teorema de Suszko

Como dissemos anteriormente, as lógicas multivaloradas foram propostas de análise de muitos problemas que a lógica clássica, bivalorada, supostamente não é adequada para lidar por ser muito forte. Por exemplo, diversas lógicas multivaloradas foram propostas para solucionar o problema da antinomia do mentiroso uma vez que a presença dessa sentença em uma teoria clássica acarreta contradições. Dentre elas, as mais conhecidas são a lógicas

trivaloradas $Ł_3$ de Łukasiewicz, B_3 de Bochvar, K_3 de Kleene e LP de Asenjo/Priest.[9] O problema dos futuros contingentes foi também objeto de análise das lógicas multivaloradas. Segundo Łukasiewicz, sentenças sobre o futuro tais como *amanhã haverá uma batalha naval* não são verdadeiras nem falsas. Portanto, o terceiro excluído não vale irrestritamente. Como uma proposta de análise, Łukasiewicz introduz o sistema $Ł_3$ onde o terceiro valor, representado numericamente por ½, é interpretado como *possível*. Łukasiewicz defende que sentenças sobre o futuro devem receber o valor ½ ao invés de 1 ou 0 por serem possíveis e não necessárias. Por último, damos como exemplo uma aplicação interessante das lógicas multivaloradas. Por volta dos anos de 1919, Bernays utiliza tabelas de verdade com quatro valores para provar a independência dos axiomas da lógica proposicional clássica. De certo modo, Bernays pode ser creditado como um dos criadores da teoria das matrizes lógicas. Entretanto, essa teoria é creditada a Łukasiewicz, Tarski e Lindenbaum pelo trabalho mais sistemático acerca das matrizes.

Como vimos, as lógicas multivaloradas possuem aplicações filosóficas e técnicas interessantes. Por outro lado, é legítimo questionar se a estipulação de mais valores de verdade além do verdadeiro e do falso é suficiente para invalidar o *Princípio de Bivalência*, segundo o qual toda sentença é verdadeira ou falsa. Suszko (SUSZKO, 1977) observa que a distinção entre valores distinguidos e não-distinguidos mostra que as lógicas multivaloradas são, na verdade, bivalentes. Como podemos ver na seção anterior, os conceitos de tautologia e consequência lógica são definidos, respectivamente, como uniformidade de valores distinguidos e preservação de valores distinguidos das premissas para a conclusão. Suszko argumenta que existe uma diferença entre *valores algébricos* e *valores lógicos*. Os valores algébricos são entendidos como referentes admissíveis de fórmulas, enquanto os valores lógicos são o *verdadeiro* e o *falso*. Tomando a lógica LP como exemplo, os valores 1, ½ e 0 são valores algébricos e o verdadeiro e o falso são os valores lógicos. Ele defende que os valores algébricos e lógicos são de natureza completamente dis-

[9] Uma boa exposição de lógicas multivaloradas pode ser encontrada em Malinowski (MALINOWSKI, 1993). Embora a lógica LP seja mais conhecida a partir dos trabalhos de Priest, ela foi elaborada por Asenjo em (ASENJO, 1966).

tinta. Em virtude desse caráter dicotômico dos valores, Suszko afirma que *toda lógica é logicamente bivalorada*. Essa última afirmação tornou-se conhecida como *Tese de Suszko.*

Suszko defende que a multiplicação de valores de verdade promovida por Łukasiewicz ocasionou na abolição do *Axioma Fregeano* (**AF**), que Suszko formula como segue:

> (**AF**) *Todas as sentenças verdadeiras* (*respectivamente, falsas*) *possuem um referente em comum.*

De acordo com (**AF**), todas as sentenças verdadeiras referem ao verdadeiro e as falsas referem ao falso. Considerando novamente LP podemos nos perguntar em que sentido ela viola esse axioma. As matrizes de LP tomam 1 e ½ como valores distinguidos e 0 como não-distinguido. Assim, nem todas as sentenças verdadeiras terão o verdadeiro como referente, pois uma sentença φ pode ter 1 ou ½ como referente. No caso da lógica proposicional clássica, (**AF**) permanece intacto, pois os referentes admissíveis das fórmulas colapsam com seus valores lógicos. E esse é exatamente o ponto no qual as lógicas multivaloradas divergem da clássica. Nesse aspecto, as fórmulas exercem um duplo valor semântico, pois nem sempre seus referentes admissíveis correspondem aos valores lógicos.

É importante esclarecer que Suszko não endossa (**AF**). Segundo Malinowski (MALINOWSKI, 1985), Suszko rejeita (**AF**) ao defender que os referentes das sentenças são *estados de coisas* em vez de valores de verdade. Os estados de coisas são entendidos de acordo com a perspectiva wittgensteiniana em (WITTGENSTEIN, 2010), que podemos interpretar como uma combinação de objetos. Por outro lado, Suszko defende que sentenças podem somente ser verdadeiras ou falsas. Em decorrência da diferença entre o que sentenças referem e quais são seus valores lógicos, Suszko defende que (**AF**) promoveu uma confusão a esse respeito. E tal confusão levou Łukasiewicz a introduzir a *possibilidade* como um valor lógico.

É importante neste ponto dizer que a Tese de Suszko por si não é suficiente para mostrar que toda lógica é logicamente bivalorada. Deve ser mostrado, ainda, que uma tal redução é possível. Embora Suszko não apresente um método construtivo para provar que toda lógica possui dois valores lógicos, no artigo (SUSZKO, 1977) ele sugere que todos os valores distinguidos sejam associados ao valor verdadeiro (T) e que todos os valores não-distinguidos sejam associados ao falso (F). Em outras palavras, dada uma matriz $\mathscr{M}_1 = (\{v_1, \ldots, v_n\}, \neg, \rightarrow, \{v_1, \ldots, v_k\})$ para $k < n$, associamos todos os valores $v_1, \ldots, v_k$ ao valor T e os demais valores $v_{k+1}, \ldots, v_n$ ao F. Mais adiante discutiremos os efeitos dessa redução à bivalência.[10]

Definição 2.3. Uma bivaloração b é uma $b: For^L \rightarrow \{T,F\}$ definida pelas seguintes cláusulas:

(i) $b(\varphi \rightarrow \psi) = T$ se e somente se $b(\neg\varphi) = T$ ou $b(\psi) = T$
(ii) $b(\varphi \rightarrow \psi) = F$ *sse* $b(\varphi) = T$ e $b(\psi) = F$
(iii) $b(\neg(\varphi \rightarrow \psi)) = T$ *sse* $b(\varphi) = T$ e $b(\neg\psi) = T$
(iv) $b(\neg\neg\varphi) = T$ *sse* $b(\varphi) = T$
(v) se $b(\neg\varphi) = T$ então $b(\varphi) = F$ ou $b(\varphi) = T$
(vi) se $b(\neg\varphi) = F$ então $b(\varphi) = T$

Dizemos que $\Gamma \vDash^2_{LP} \varphi$ se e somente se toda bivaloração b tal que $b(\gamma) = T$, para todo $\gamma \in \Gamma$ implica $b(\varphi) = T$.

Antes de prosseguirmos, é importante esclarecer que o fato de haver uma semântica bivalorada para uma lógica multivalorada não constitui um paradoxo. Como observam Caleiro *et al* (CALEIRO *et al*, 2007), embora pareça estranho ouvir que uma lógica multivalorada **L** é bivalorada, é necessário perceber que existe uma troca. A semântica multivalorada **L** é verofuncional. Ou seja, o valor de verdade de φ é determinado pelos valores de verdade de suas

[10] Como dissemos em uma nota de rodapé anterior, estamos trabalhando em uma perspectiva totalmente semântica. Os resultados nos quais Suszko se baseia são provenientes da lógica abstrata, que basicamente considera uma lógica **L** como um par $(L, \Vdash)$, onde L é uma linguagem e $\Vdash$ é uma relação de consequência definida a partir das fórmulas de L. Aqui, por comodidade, adotaremos uma perspectiva totalmente semântica, estando cientes de que nossa construção é possível em decorrência de resultados mais gerais acerca de lógicas abstratas.

subfórmulas. No caso da semântica bivalorada para LP é fácil ver que a verdade de ¬φ não é determinado pela falsidade de φ. Ou seja, *b* não é verofuncional. Portanto, a transição de uma semântica verofuncional para uma semântica bivalorada para uma lógica multivalorada **L** tem como preço a perda da verofuncionalidade. Mostraremos agora que a semântica definida na Definição 2.3 é equivalente à semântica multivalorada de LP.

Teorema 2.4. Dada uma bivaloração b: $For^L \rightarrow \{T,F\}$, existe uma valoração v: $For^L \rightarrow \{1,½,0\}$ tal que

$v(\varphi) \in \{1,½\}$ sse $b(\varphi) = T$

Prova. Definimos *v* como segue:

(a) $v(\varphi) = 1$ se $b(\varphi) = T$ e $b(\neg\varphi) = F$
(b) $v(\varphi) = ½$ se $b(\varphi) = T$ e $b(\neg\varphi) = T$
(c) $v(\varphi) = 0$ se $b(\varphi) = F$ e $b(\neg\varphi) = T$

Provaremos que *v* é uma valoração de LP. O caso atômico é imediato. Mostraremos agora o caso dos demais conectivos booleanos.

1) $\varphi = \neg\psi$
Se $v(\psi) = 1$, então por definição $b(\psi) = T$ e $b(\neg\psi) = F$. Pela definição de *b*, obtemos $b(\neg\neg\psi) = T$. Uma vez que $b(\neg\psi) = F$ e $b(\neg\neg\psi) = T$, então $v(\neg\psi) = 0$.
Se $v(\psi) = ½$, então por definição $b(\psi) = T$ e $b(\neg\psi) = T$. Da definição de *b*, obtemos $b(\neg\neg\psi) = T$, $b(\neg\neg\neg\psi) = T$. Logo, $v(\neg\psi) = ½$.
Se $v(\psi) = 0$, então por definição $b(\psi) = F$ e $b(\neg\psi) = T$. Pela definição de *b*, obtemos $b(\neg\neg\psi) = F$ e $b(\neg\neg\neg\psi) = T$. Daqui, $v(\neg\psi) = 1$.
2) $\varphi = \psi \rightarrow \gamma$

(2.1) Se $v(\psi) = 1$ e $v(\gamma) = 1$, obtemos por definição ($b(\psi) = T$ e $b(\neg\psi) = F$) e ($b(\gamma) = T$ e $b(\neg\gamma) = F$). Então $b(\psi \rightarrow \gamma) = T$ e $b(\neg(\psi \rightarrow \gamma)) = F$. Logo $v(\psi \rightarrow \gamma) = 1$.

(2.2) Se $v(\psi) = 1$ e $v(\gamma) = ½$, obtemos por definição ($b(\psi) = T$ e $b(\neg\psi) = F$) e ($b(\gamma) = T$ e $b(\neg\gamma) = T$). Então $b(\psi \rightarrow \gamma) = T$ e $b(\neg(\psi \rightarrow \gamma)) = T$. Logo $v(\psi \rightarrow \gamma) = ½$.

(2.3) Se $v(\psi) = 1$ e $v(\gamma) = 0$, obtemos por definição ($b(\psi) = T$ e $b(\neg\psi) = F$) e ($b(\gamma) = F$ e $b(\neg\gamma) = T$). Então $b(\psi \rightarrow \gamma) = F$ e $b(\neg(\psi \rightarrow \gamma)) = T$. Logo $v(\psi \rightarrow \gamma) = 0$.

(2.4) Se $v(\psi) = ½$ e $v(\gamma) = 1$, obtemos por definição ($b(\psi) = T$ e $b(\neg\psi) = T$) e ($b(\gamma) = T$ e $b(\neg\gamma) = F$). Então $b(\psi \rightarrow \gamma) = T$ e $b(\neg(\psi \rightarrow \gamma)) = F$. Logo $v(\psi \rightarrow \gamma) = 1$.

(2.5) Se $v(\psi) = ½$ e $v(\gamma) = ½$, obtemos por definição ($b(\psi) = T$ e $b(\neg\psi) = T$) e ($b(\gamma) = T$ e $b(\neg\gamma) = T$). Então $b(\psi \rightarrow \gamma) = T$ e $b(\neg(\psi \rightarrow \gamma)) = T$. Logo $v(\psi \rightarrow \gamma) = ½$.

(2.6) Se $v(\psi) = ½$ e $v(\gamma) = 0$, obtemos por definição ($b(\psi) = T$ e $b(\neg\psi) = T$) e ($b(\gamma) = F$ e $b(\neg\gamma) = T$). Então $b(\psi \rightarrow \gamma) = T$ e $b(\neg(\psi \rightarrow \gamma)) = T$. Logo $v(\psi \rightarrow \gamma) = ½$.

Os casos (2.7) – (2.9) são os casos em que $v(\psi) = 0$. Da definição de b, é imediato que $b(\neg\psi) = T$. Claramente $b(\psi \rightarrow \gamma) = T$ e $b(\neg(\psi \rightarrow \gamma)) = F$. Então $v(\psi \rightarrow \gamma) = 1$. Isso conclui a demonstração. *Q.E.D.*

Teorema 2.5. Dada uma valoração $v \in sem_{LP}$, existe uma bivaloração $b: For^{L} \rightarrow \{T,F\}$ tal que:

$b(\varphi) = T$ sse $v(\varphi) \in \{1,½\}$, para toda $\varphi \in For^{L}$.

Prova. A prova é por indução na complexidade de φ.

1) $\varphi = p \in At$

$b(p) = T$ sse $v(p) = 1$ ou $v(p) = ½$ sse $v(p) \in \{1,½\}$.

2) $\varphi = \neg\psi$ e $\psi = p$

$b(\neg p) = T$ sse $b(p) = F$ ou $b(p) = T$ sse $v(p) = 0$ ou $v(p) \in \{1,½\}$ sse $v(\neg p) \in \{1,½\}$.

2.1) $\varphi = \neg\psi$ e $\psi = \neg\gamma$

$b(\neg\neg\gamma) = T$ sse $b(\gamma) = T$ sse (H.I.) $v(\gamma) \in \{1,½\}$ sse $v(\neg\neg\gamma) \in \{1,½\}$.[11]

2.2) $\varphi = \neg\psi$ e $\psi = \gamma \rightarrow \beta$

$b(\neg(\gamma \rightarrow \beta)) = T$ sse $b(\gamma) = T$ e $b(\neg\beta) = T$ sse $b(\gamma) = T$ e ($b(\beta) = F$ ou $b(\beta) = T$) sse (H.I.) $v(\gamma) \in \{1,½\}$ e ($v(\beta) = 0$ ou $v(\beta) \in \{1,½\}$) sse $v(\gamma \rightarrow \beta) \in \{0,½\}$ sse $v(\neg(\gamma \rightarrow \beta)) \in \{1,½\}$.

3) $\varphi = \gamma \rightarrow \beta$

$b(\gamma \rightarrow \beta) = T$ sse $b(\neg\gamma) = T$ ou $b(\beta) = T$ sse ($b(\gamma) = F$ ou $b(\gamma) = T$) ou $b(\beta) = T$ sse (H.I.) ($v(\gamma) = 0$ ou $v(\gamma) \in \{1,½\}$) ou $v(\beta) \in \{1,½\}$ sse $v(\gamma \rightarrow \beta) \in \{1,½\}$. Isso conclui o que pretendemos demonstrar. *Q.E.D.*

Os Teoremas 2.4 e 2.5 mostram que LP pode ser apresentada a partir de uma semântica bivalente. Como Caleiro *et al* (CALEIRO, et al, 2007) observam, esse resultado redutivo de Suszko depende da estruturalidade da relação de consequência lógica. Por outro lado, existem reduções mais gerais que não pressupõem estruturalidade. Um exemplo interessante desta última observação são as semânticas bivaloradas de da Costa (DA COSTA, 1974) para a hierarquia de lógicas C_n, que em geral não possuem semântica multivalorada finita. É importante dizer que as bivalorações de da Costa foram descobertas independentemente das de Suszko. E o fato de não pressupor a estruturalidade da relação de consequência torna os resultados de da Costa mais gerais que os de Suszko. No que diz respeito à construtividade do resultado, Caleiro *et al* (CALEIRO *et al*, 2007) observam que o resultado de Suszko não é construtivo no sentido que Suszko não apresenta um método construtivo que permita obter uma semântica bivalorada para cada lógica multivalorada. Além disso, por mais que o resultado segundo o qual toda lógica multivalorada é logicamente bivalorada seja atribuído a Suszko, a prova deste fato encontra-se no livro de Malinowski (MALINOWSKI, 1993). Em (CALEIRO *et al*, 2007)

[11] Hipótese de Indução.

encontramos um método construtivo que nos permite obter, para cada lógica multivalorada **L**, uma semântica bivalente que caracteriza **L**.

As semânticas do estilo apresentado na Definição 2.3 são frequentemente criticadas no que diz respeito a sua intuitividade. Por exemplo, ela não explica o processo que nos permitiu obter $b(\varphi) = b(\neg\varphi) = \mathrm{T}$. No caso da semântica matricial a resposta é imediata, pois $v(\varphi) = v(\neg\varphi)$ quando e somente quando $v(\varphi) = ½$. Por outro lado, é possível obter semânticas bivalentes que sejam capazes de explicar o significado das constantes lógicas da linguagem lógica. Por exemplo, Carnielli Lima Marques (CARNIELLI & LIMA MARQUES, 1999) apresentam as chamadas *Semânticas de Sociedades*, que são semânticas bivaloradas para lógicas multivaloradas capazes de descrever situações onde dois ou mais agentes diferentes podem aceitar proposições contraditórias, e situações onde dois ou mais agentes diferentes podem rejeitar tanto uma proposição quanto sua contraditória. Em (BEZERRA, 2017) pode ser encontrado tal semântica para LP. A razão pela qual apresento a semântica da Definição 2.3 em detrimento de outra deve-se pelo fato de que meu objetivo é mostrar a simples possibilidade de apresentar uma semântica bivalente para uma lógica multivalorada.

Embora a verificação de tautologias em lógicas caracterizadas por semânticas bivaloradas não seja tão simples quanto a verificação por meio de tabelas de verdade, Caleiro *et al* (CALEIRO *et al*, 2010) defendem que essas semânticas possuem aplicações interessantes no estudo de sistemas formais dentre as quais destacamos: (i) tais semânticas dão um procedimento de decisão para lógicas que não possuem semântica finitamente valorada, como é o caso da hierarquia C_n anteriormente citadas, bem como outros sistemas lógicos paraconsistentes; (ii) elas facilitam a construção de sistemas de prova, tal como os tablôs semânticos para essas lógicas, o que permite um procedimento uniforme de provas de completude. Essas aplicações mostram que, além de ser uma tese filosófica acerca da multivaloração, a Tese de Suszko mostra-se como uma ferramenta interessante de investigação lógica.

da Costa *et al* (DA COSTA *et al*, 1996) levantam críticas à Tese de Suszko tendo como base a multivaloração inferencial de Malinowski (MALI-

NOWSKI, 1994). Malinowski reconhece que a divisão bipartite do conjunto de valores lógicos implica que a relação de consequência $\vDash$ seja tarskiana estrutural e, portanto, bivalente. Contudo, ele afirma que a multivaloração lógica ainda é possível da seguinte maneira:

> The departure is a division of the matrix universe into three subsets of: *rejected* elements, *accepted* elements and all other elements. On such grounds it was possible to define the relation being a formal counterpart of reasoning admitting rules of inference which from non-rejected assumptions lead to accepted conclusions (...) The relation was then called, somewhat inaccurately, a *q-consequence* (...) According to this φ is inferred from the set of premisses *X*, whenever it is the case that if all premisses are not rejected then φ is accepted. (MALINOWSKI, 2007, p. 68)

Em que sentido podemos dizer que a definição apresentada por Malinowski é um problema para a Tese de Suszko? Essa definição não configura um contraexemplo ao resultado redutivo de Suszko, pois Tsuji (TSUJI, 1998) mostra que a relação de consequência definida por Malinowski não é reflexiva. Portanto, não é tarskiana estrutural e não configura um contraexemplo ao teorema de Suszko. Mesmo assim, da Costa *et al* argumentam que ela configura um contraexemplo à Tese de Suszko, uma vez que é possível mostrar que a *q-consequência* admite trivalorações lógicas. De fato, Malinowski (MALINOWSKI, 1990) mostra que tais relações de consequência não admitem necessariamente redução a um modelo bivalente. Em virtude dessa possibilidade, da Costa *et al* argumentam que não temos razão *a priori* para defender tal tese.

Embora possa reconhecer certo mérito na argumentação de da Costa *et al*, reconheço que $\vDash_{\mathbf{L}}$ e a q-consequência capturam diferentes aspectos do conceito de consequência lógica. Como podemos ver na citação acima, a q-consequência não é uma relação que preserva valores designados do conjunto de premissas para a conclusão, mas que preserva atitudes proposicionais. Suponha a possibilidade de definir uma variante da q-consequência para o caso de LP de maneira que o valor ½ tanto ao conjunto de valores aceitos quanto ao conjunto de valores rejeitados. Chamemos tal relação de $\vDash_{\text{d-LP}}$. Claramente $\vDash_{\text{LP}}$ e $\vDash_{\text{d-LP}}$ capturam diferentes aspectos de LP, podendo ser utilizada

para capturar aspectos hipotéticos das ciências empíricas.[12] Consequentemente, $\vDash_{\text{d-LP}}$ perderá ao menos uma das propriedades de $\vDash_{\text{LP}}$. Sob essa ótica não é claro se LP munida de $\vDash_{\text{d-LP}}$ captura o aspecto de preservação de verdade que $\vDash_{\text{LP}}$ captura. Em suma, $\vDash_{LP}$ e $\vDash_{\text{d-LP}}$ não descrevem as mesmas coisas. A primeira descreve contexto de preservação de verdade enquanto o segundo descreve contextos voltados para hipóteses de ciência empírica. Então, não podemos dizer que uma relação de consequência do tipo $\vDash_{\text{d-LP}}$ é um contraexemplo à Tese de Suszko, que é uma tese alética. A relação de consequência proposta por Malinowski captura aspectos puramente inferenciais de uma lógica. O que a objeção de da Costa *et al* mostra é que lógicas, de maneira geral, não possuem interpretações canônicas.[13]

O que a Tese de Suszko implica é que, de uma perspectiva alética, a lógica é bivalorada. Aqui não nos comprometeremos com teses metafísicas que a suportam. De um ponto de vista ontológico, pode ser argumentado que existem muitos valores de verdade além do verdadeiro e falso. Por outro lado, uma resposta definitiva a esse problema pode não ser possível. Adotarei uma justificação pragmática a seu favor. O fato de lógicas não-clássicas terem uma metateoria clássica já é um indício em defesa do valor pragmático da Tese de Suszko.[14] Além disso, a própria definição de consequência lógica adotada por Suszko em (SUSZKO, 1977) captura intuições importantes presentes na prática argumentativa em matemática. Neste ponto, estou de acordo com Beall & Restall (BEALL & RESTALL, 2006) que uma relação de consequência, se entendida como preservadora de *verdade* do conjunto de premissas para a conclusão, intuitivamente satisfaz as propriedades da relação de consequência

[12] Para uma exposição sistemática de relações generalizadas de consequência, convidamos o leitor a consultar a tese de doutoramento de Carolina Blasio (BLASIO, 2017).

[13] A tese segundo a qual uma lógica não possui interpretação canônica não é nova na literatura. Podemos encontrá-la, por exemplo, em (BARRIO, 2018). Nesse artigo, Barrio concentra-se em lógicas paraconsistentes. No presente trabalho me concentro em lógicas multivaloradas, de maneira geral.

[14] Em (BACON, 2013), Bacon argumenta em favor da possibilidade de definir uma metateoria não-clássica para uma lógica não-clássica. Defendemos, entretanto, que essa possibilidade não constitui necessariamente um contraexemplo para a Tese de Suszko, visto que em muitas lógicas o conceito de verdade pode não possuir papel fundamental, como é o caso de certas lógicas paraconsistentes defendidas por Carnielli & Rodrigues (CARNIELLI & RODRIGUES, 2019).

tarskiana. Considere, por exemplo, a seguinte definição informal de consequência lógica:

> (*) *A é consequência lógica de Σ se e somente se em todo caso que as sentenças de Σ são verdadeiras, A é verdadeira.*

Uma definição no estilo (*) pode ser encontrada em (BEALL & RESTALL, 2006). A sentença A e o conjunto Σ de sentenças estão sendo consideradas como pertencendo à linguagem natural e sendo livres de ambiguidade. A informalidade de (*) deve-se a não especificação da palavra 'caso'. Pode ser facilmente mostrado que a definição (*) satisfaz as propriedades da relação de consequência $\vDash_{\mathbf{L}}$ apresentadas acima. Então, a relação $\vDash_{\mathbf{L}}$ consegue capturar uma noção importante de consequência lógica na linguagem natural. Mas, o fato da relação $\vDash_{\mathbf{L}}$ capturar uma noção de consequência lógica na linguagem natural não significa que ela captura uma noção intuitiva, pré-teórica, de consequência lógica. Como Smith (SMITH, 2011) argumenta, por mais que $\vDash_{\mathbf{L}}$ consiga capturar uma noção *informal* de consequência lógica, isso não significa dizer que ela captura uma noção *pré-teórica/primitiva* de consequência lógica. Uma noção tal como (*) é resultado de um refinamento teórico necessário para dizermos que ela se aproxima de uma noção formal de consequência teoricamente relevante como $\vDash_{\mathbf{L}}$. Consequentemente, não podemos dizer que (*) é *a* noção primitiva de consequência lógica. Isso não constitui um problema visto que nossas noções pré-teóricas podem ser confusas e, até mesmo, errôneas. Como Smith argumenta, é muito difícil precisar qual é a noção pré-teórica de consequência. Possivelmente, ela carrega elementos de relevância e não de preservação de verdade. Mas não sabemos. O que podemos realmente dizer acerca de (*) é que ela consegue descrever uma relação importante de consequência lógica que pode ser usada tanto na linguagem natural quanto no raciocínio matemático. Nesse aspecto, podemos defender a razoabilidade da relação de consequência que sustenta a própria Tese de Suskzo.

Mesmo as lógicas que não são redutíveis a modelos bivalorados não parecem contradizer a Tese de Suzsko um ponto de vista filosófico, pois pode ser argumentado que suas relações de consequência não preservam verdade no

sentido estrito do termo. Podem, por exemplo, interpretar os valores algébricos como atitudes proposicionais tais como 'aceitação' e 'rejeição' em relação às variáveis proposicionais. Essa pluralidade de relações de consequência correspondentes a uma mesma linguagem não constitui um problema, pois corrobora com a tese segundo a qual uma lógica não possui uma interpretação canônica (BARRIO, 2018). Enquanto sistema formal, uma lógica pode admitir muitas interpretações. Os modelos que apresentamos para LP corrobora com essa afirmação. Mostramos ser possível apresentar uma semântica matricial trivalorada, bem como uma semântica de bivalorações para LP.

2.2. É a Tese de Suszko um golpe definitivo contra as lógicas multivaloradas?

Se a Tese de Suszko é razoável, poderíamos nos perguntar se ela é um obstáculo para lógicas multivaloradas. Defendo que essas lógicas podem conviver com a tese segundo a qual existem somente dois valores lógicos. Tal coexistência é possível em virtude do fato de essas lógicas, enquanto sistemas formais, não se reduzirem às semânticas de matrizes. Argumentei anteriormente que a possibilidade de apresentar mais de uma semântica para lógica mostra que não existe uma interpretação canônica para uma lógica. A razão de chamarmos uma lógica **L** de *n*-valorada deve-se ao fato de *n* ser a menor cardinalidade do conjunto de valores de verdade das matrizes que caracterizam **L**. Por outro lado, não podemos reduzir **L** à semântica de matrizes. Nesse sentido tais semânticas podem ser vistas como possíveis representações para essas lógicas. Segundo Marcos (MARCOS, 2009), a não ser que exista um resultado que impeça que uma lógica tenha uma determinada propriedade, é um erro classificá-la tendo em vista a semântica a partir da qual essa lógica é *circunstancialmente* apresentada. Ou seja, não podemos dizer, do ponto de vista semântico, que uma lógica tem essencialmente uma determinada propriedade já que é a princípio possível dar uma semântica para essa lógica que não tenha essa propriedade, a não ser que a lógica em questão falhe em possuir determinada propriedade.[15] Por exemplo, não podemos dizer que a semântica

[15] A lógica intuicionista é um exemplo de lógica que não pode ser apresentada por meio de uma semântica matricial finita. Tal impossibilidade foi demonstrada por Gödel em (GÖ-

matricial de LP nos compromete com um terceiro valor de verdade além do verdadeiro e do falso já que é possível caracterizar essa lógica por meio de uma semântica bivalente na qual esse terceiro valor não existe. Assim, por mais que essas lógicas sejam apresentadas através de uma semântica matricial que admitem mais valores (algébricos) de verdade além do verdadeiro e do falso, elas podem ser representadas por semânticas bivalentes. E elas admitem semânticas bivalentes devido aos resultados redutivos aos quais estão sujeitas. Nesse aspecto, cada semântica se apresenta como uma maneira de explicar o vocabulário de uma lógica **L**. Do ponto de vista das bivalorações, por exemplo, podemos entender quais são os princípios válidos sobre verdade e falsidade em **L**. Por essas razões, defendo que as lógicas multivaloradas podem coexistir com a Tese de Suszko.

3. Pluralismo lógico e Tese de Suszko

O pluralismo lógico é uma posição filosófica que reconhece a existência de mais de uma lógica como correta. O debate recente acerca do pluralismo lógico ganhou força a partir de uma série de trabalhos de Beall & Restall, principalmente com o livro (BEALL & RESTALL, 2006). Como apresentamos anteriormente, eles propõem a definição informal (*) de consequência lógica que é compatível com mais de uma especificação da palavra 'caso'. Se, por exemplo, substituirmos a 'caso' por 'construções', obtemos a lógica intuicionista. Por outro lado, a substituição da palavra 'caso' por 'modelos,' entendidos como modelos tarskianos para lógica de primeira ordem, obtemos a própria lógica de primeira ordem. Ou seja, diferentes especificações da palavra 'caso' resultam em definições formais de consequência.[16]

Entretanto, uma objeção a qual muitas versões do pluralismo estão sujeitas, inclusive a formulada por Beall & Restall, é a *objeção da mudança de significado*. Essa objeção foi formulada por Quine (QUINE, 1986). De acordo

DEL, 1932).

[16] É importante ressaltar que nem toda especificação da palavra 'caso' define uma relação de consequência. Em (BEALL & RESTALL, 2006), Beall & Restall defendem que a instância resultante da substituição deve satisfazer alguns critérios (*formalidade*, *necessidade* e *normatividade*) para ser considerada como uma instância de uma relação de consequência. Para mais detalhes, convidamos o leitor a conferir o referido livro.

com Quine, proponentes de sistemas lógicos substancialmente distintos falam sobre coisas distintas. Por sistemas lógicos substancialmente distintos entendo sistemas que discordam em ao menos um princípio lógico. Sejam $\mathbf{L}_1$ e $\mathbf{L}_2$ dois sistemas lógicos substancialmente distintos. Consequentemente, ainda que $\mathbf{L}_1$ e $\mathbf{L}_2$ possuam ao menos um conectivo *c* em comum, o significado de *c* em $\mathbf{L}_1$ diferirá do significado de *c* em $\mathbf{L}_2$.[17]

A versão do pluralismo com o qual me ocuparei aqui é o pluralismo de relações de consequência proposto por Hjortland (HJORTLAND, 2011). No referido artigo, Hjortland propõe uma versão do pluralismo lógico que ele chama de *pluralismo intra-teórico*. O pluralismo intra-teórico consiste na presença de mais uma relação de consequência dentro de uma mesma teoria lógica, sendo esta entendida como um par $(L, \Vdash)$, onde *L* é uma linguagem e $\Vdash$ é uma relação de consequência definida em *L*. Ou seja, Hjortland admite teorias lógicas do tipo $(L, \Vdash_1, ..., \Vdash_n)$, onde $\Vdash_1, ..., \Vdash_n$ denotam relações de consequência distintas definidas na mesma linguagem *L*, ao mesmo tempo em que admite que o significado dos conectivos seja fixo em cada uma das $\Vdash_1, ..., \Vdash_n$. Do ponto de vista do cálculo de sequentes, isso significa que as regras dos conectivos são mantidas fixas em todas as relações de consequência, ao passo que as regras estruturais dos sequentes podem variar em cada uma dessas relações. Isso significa que o núcleo de cada conectivo é preservado nas diferentes relações de consequência $\Vdash_i$, para $1 \leq i \leq n$.[18]

Já do ponto de vista semântico, com o qual irei me concentrar aqui, isso significa que o significado dos conectivos é dado pela tabela dos valores algébricos. Então, as tabelas dos valores algébricos são mantidas fixas, ao passo que a variação consiste na escolha dos valores distinguidos em cada uma das relações $\vDash_i$, para $1 \leq i \leq n$. Ou seja, a matriz $\mathscr{M}_i$ de cada $\vDash_i$ possuirá o mesmo conjunto de valores algébricos, e as operações possuirão as mesmas tabelas. O que muda, portanto, é o conjunto *D* de valores distinguidos. Essa versão do pluralismo aplica-se a todas às lógicas multivaloradas.[19]

---

[17] É importante ter em mente que a Quine considera a lógica clássica como lógica ortodoxa, a lógica do discurso científico. A questão da mudança de assunto direciona-se principalmente às lógicas heterodoxas, isto é, às lógicas divergentes da lógica clássica.

[18] Para os detalhes sintáticos, convidamos o leitor a conferir (HJORTLAND, 2011).

Considere novamente o exemplo da lógica LP.[20] Como apresentado anteriormente, sua matriz $\mathscr{M}$ possui o conjunto {1,½} de valores distinguidos. É sabido que a lógica K_3 (RESCHER, 1969) é caracterizada pela matriz $\mathscr{M}$'= ({1,½,0}, ¬, →, {1}), onde os operadores ¬ e → possuem as mesmas tabelas de valores algébricos que os operadores da matriz $\mathscr{M}$ de LP. Ou seja, assumindo que o significado de um conectivo possa ser dado pela tabela de valores algébricos que o caracteriza, obtém-se que os conectivos de LP e K_3 possuem o mesmo significado. A diferença entre as lógicas reside nas relações de consequência que as caracterizam. Portanto, das relações de consequência $\vDash_{LP}$ e $\vDash_{K3}$ definidas a partir das matrizes $\mathscr{M}$ e $\mathscr{M}$' obtemos uma teoria (L,$\vDash_{LP}$, $\vDash_{K3}$), onde L é a linguagem comum a LP e K_3.

Ocuparei-me agora com a questão de como a Tese de Suszko se relaciona com o pluralismo intra-teórico. Como vimos na subseção 2.2, os Teoremas 2.4 e 2.5 mostram que $\vDash_{LP} = \vDash^2_{LP}$. É possível tanto definir uma semântica nos moldes da Definição 2.3 quanto provar, utilizando um raciocínio similar, dois teoremas análogos aos Teoremas 2.4 e 2.5 em relação à lógica K_3. Ou seja, é possível provar que $\vDash_{K3} = \vDash^2_{K3}$. Haverá, por outro lado, ao menos uma diferença entre as bivalorações de LP e as bivalorações de K_3, uma vez que LP é paraconsistente e K_3 é paracompleta, isto é, que não valida o princípio do terceiro excluído.

Se o que muda nas matrizes é a escolha dos elementos do conjunto de valores algébricos que tomamos como distinguidos, consequentemente, o que muda nas bivalorações é a escolha da razão da verofuncionalidade da negação. Nas bivalorações de LP, a verdade de ¬φ não é determinada pela falsidade de φ. Por outro lado, nas bivalorações de K_3, a falsidade de ¬φ não é determina-

[19] Uma maneira de introduzir a discussão do pluralismo lógico com respeito às lógicas multivaloradas seria articular a relação entre valores algébricos e lógicas instrumentais, por exemplo. Na literatura podemos encontrar diversos exemplos interessantes dessa relação. A própria lógica K_3 é um exemplo interessante. A proposta de Kleene visava descrever funções recursivas parciais que podem estar indefinidas para determinados objetos em um dado domínio. Por outro lado, a proposta de Kleene não visava apresentar K_3 como uma lógica que caracteriza princípios mais gerais do raciocínio. Ou seja, K_3 não foi pretendida como uma *lógica canônica* (PRIEST, 2006b). Os valores intermediários de lógicas multivaloradas, tais como as *nonsense-logics* (PIRÓG-RZEPECKA, 1977) foram introduzidos como instrumentos para caracterizar certos aspectos de raciocínio que a lógica clássica não trata, ao menos imediatamente. Essa discussão, embora interessante e importante de ser feita, não será abordada em detalhes neste artigo.

[20] Essa mesma lógica é objeto de análise no artigo de Hjortland.

da pela verdade de φ. De uma perspectiva puramente técnica, as bivalorações explicitam as condições de uma fórmula ter um valor (não-) designado. Seja *D* o conjunto de valores designados.

(i) $v(\varphi \rightarrow \psi) \in D$ se e somente se $v(\neg\varphi) \in D$ ou $v(\psi) \in D$
(ii) $v(\varphi \rightarrow \psi) \notin D$ *sse* $v(\varphi) \in D$ e $v(\psi) \notin D$
(iii) $v(\neg(\varphi \rightarrow \psi)) \in D$ *sse* $v(\varphi) \in D$ e $v(\neg\psi) \in D$
(iv) $v(\neg\neg\varphi) \in D$ *sse* $v(\varphi) \in D$
(v) se $v(\neg\varphi) \in D$ então $v(\varphi) \notin D$ ou $v(\varphi) \in D$
(vi) se $v(\neg\varphi) \notin D$ então $v(\varphi) \in D$

Já no caso K_3, as cláusulas (v) e (vi) são modificadas como segue

(v') se $v(\neg\varphi) \notin D$ então $v(\varphi) \notin D$ ou $v(\varphi) \in D$
(vi') se $v(\neg\varphi) \in D$ então $v(\varphi) \notin D$

As cláusulas acima são descrições mais gerais do que ocorre nas tabelas de verdade de LP e de K_3. Os conectivos de ambas as lógicas, portanto, possuem comportamentos similares em cenários distintos.

Note que essa argumentação mostra a compatibilidade do pluralismo intra-teórico com a Tese de Suszko, uma vez que as bivalorações podem ser vistas como explicitando o que apresentamos nos dois últimos parágrafos. Além disso, não parece haver inconsistência entre a Tese de Suszko e a afirmação de que o significado dos conectivos é dado pelas suas tabelas de valores algébricos, uma vez que a semântica de bivalorações e a semântica matricial capturam diferentes aspectos de uma mesma lógica.

Em última instância, a relação entre a pluralidade de lógicas e a Tese de Suszko mostra uma independência do *PNC, PTE* e *PE* em relação ao princípio de bivalência. A lógica clássica satisfaz todos esses princípios mencionados, o que pode sugerir que eles estão necessariamente interligados. Por outro lado, outras lógicas mais fracas mostram que podemos não ter todos eles conjuntamente e elas possuem legitimidade. Por exemplo, a lógica LP valida

PTE, PNC mas não valida *PE*. Já a lógica trivalorada K_3 valida *PE*, mas não valida tanto *PNC* quanto *PTE*. Outras lógicas multivaloradas exemplificam que esses três princípios são realmente independentes uns dos outros. A lógica I^l valida *PE* e *PNC*, mas não valida *PTE*. Por último, a lógica P^l valida *PTE,* mas não valida *PNC* e *PE*.[21] Diferentes lógicas podem nos mostrar maneiras segundo as quais esses princípios podem relacionar-se.

4. Considerações finais

Na apresentação das bivalorações para LP, foi dito que as semânticas do estilo da Definição 2.3 carecem de intuitividade por não oferecerem uma explicação do significado das constantes lógicas. Por outro lado, podemos entendê-las como estabelecendo normas para a verdade das fórmulas da linguagem de uma determinada lógica. Mencionei a possibilidade apresentar semânticas bivaloradas que são mais intuitivas.[22] A possibilidade de apresentar semânticas alternativas à semântica matricial permite uma perspectiva interessante às lógicas multivaloradas, que também são frequentemente criticadas por não oferecerem explicações adequadas dos seus valores intermediários. Assim, a variedade de semânticas bivaloradas para essas lógicas nos permite enxergá-las a partir de diferentes perspectivas, abrindo um horizonte de possíveis aplicações.

A Tese de Suszko pode ser vista como uma tese acerca da natureza da lógica, uma vez que ela compreende a lógica como sendo uma teoria da consequência preservadora de verdade. Vimos, contudo, que uma lógica enquanto sistema formal é aberta a diferentes interpretações e que ela não possui interpretação canônica. Por outro lado, defendo que ela é uma tese plausível caso entendamos lógica como sendo interpretada em termos aléticos. E sua compatibilidade com a tese do pluralismo lógico sugere que os próprios valores de

[21] Convidamos o leitor a checar o artigo (CARNIELLI & RODRIGUES, 1999) para uma apresentação das lógicas P^l *e* I^l.

[22] Uma semântica bivalorada que sequer mencionamos aqui é a semântica de *supervalorações*, que são capazes de oferecer uma explicação interessante das constantes lógicas. Uma aplicação bem sucedida dessa semântica foi na teoria de Kripke dos pontos fixos como uma solução ao problema dos paradoxos semânticos. Convidamos o leitor a consultar (PRIEST, 2008) para uma apresentação dessa semântica.

verdade são independentes de *PTE*, *PNC* e *PE.* Tal independência pode ajudar na compreensão do que são os valores lógicos.

A Tese de Suszko tal como é apresentada é uma tese semântica, uma vez que é enunciada em termos de conceitos puramente semânticos. Consequentemente, a redução à bivaloração concerne à teoria de modelos, foi apresentada uma teoria de modelos para LP que possui somente dois valores que coincide com as matrizes para LP no que diz respeito a suas validades. Uma vez que essas lógicas possuem um sistema dedutivo em relação aos quais é possível demonstrar teoremas de correção e completude, podemos nos perguntar se é possível formular essa tese em termos de conceitos de teoria da prova. Conjecturo que tal formulação é possível se nos atentarmos ao tablôs analíticos para lógicas multivaloradas. Carnielli (CARNIELLI, 1987) apresenta um método para construir sistemas de tablôs para qualquer lógica finitamente valorada $\mathbf{L}_n$. Esses sistemas são rotulados, no sentido de possuírem rótulos que precedem as fórmulas em cada nó do tablô. Além disso, dada uma lógica $\mathbf{L}_n$ seu sistema correspondente de tablôs terá *n* rótulos. Interessantemente, (CALEIRO *et al*, 2007) apresenta um método construtivo para obter tablôs que possuem somente dois rótulos para lógicas $\mathbf{L}_n$. Caso seja possível apresentar um método redutivo que nos permite obter um sistema de tablôs birrotulados a partir de um sistema de tablôs *n*-rotulados, teremos então uma redução *à la* Suszko de uma perspectiva sintática. A partir de um procedimento desses será possível formular uma tese filosófica mais geral que a Tese de Suszko, visto que esta última é uma tese particularmente semântica.

Referências

ASENJO, Florencio G. A calculus of antinomies. **Notre Dame Journal of Formal Logic**, v. 7, n. 1, p. 103-105, 1966.

BACON, Andrew. Non-classical metatheory for non-classical logics. **Journal of Philosophical Logic**, v. 42, n. 2, p. 335-355, 2013.

BARRIO, Eduardo Alejandro. Models & Proofs: LFIs Without a Canonical Interpretations. **Principia: an international journal of epistemology**, v. 22, n. 1, p. 87-112, 2018.

BEALL, Jeffrey C.; RESTALL, Greg. **Logical pluralism**. Oxford University Press on Demand, 2006.

BEZERRA, Edson V. **Uma análise das bivalorações do ponto de vista das semânticas de sociedade**. Dissertação de mestrado. Universidade Estadual de Campinas, 2017.

BLASIO, Carolina. **Sobre noções de consequência generalizadas e lógicas plurivalentes**. Tese de doutorado. Universidade Estadual de Campinas. 2017.

CALEIRO, Carlos; CARNIELLI, Walter A.; CONIGLIO, Marcelo; MARCOS, João. Two's company: "The humbug of many logical values". In: **Logica universalis**. Birkhäuser Basel, 2005. p. 169-189.

CALEIRO, Carlos; MARCOS, Joao. Two Many Values: An Algorithmic Outlook on Suszko's Thesis. In: **2010 40th IEEE International Symposium on Multiple-Valued Logic**. IEEE, 2010. p. 93-97.

CARNIELLI, Walter A. Systematization of finite many-valued logics through the method of tableaux. **The Journal of Symbolic Logic**, v. 52, n. 2, p. 473-493, 1987.

CARNIELLI, Walter A. Paul Bernays and the eve of non-standard models in logic. **Universal logic: an anthology**, 2012.

CARNIELLI, Walter A.; LIMA-MARQUES, Mamede. Society semantics and multiple-valued logics. **Advances in Contemporary Logic and Computer Science**, v. 235, p. 33-52, 1999.

CARNIELLI, Walter; RODRIGUES, Abilio. An epistemic approach to paraconsistency: a logic of evidence and truth. **Synthese**, v. 196, n. 9, p. 3789-3813, 2019.

DA COSTA, Newton CA et al. Malinowski and Suszko on many-valued logics: on the reduction of many-valuedness to two-valuedness. **Modern Logic**, v. 6, n. 3, p. 272-299, 1996.

DA COSTA, Newton CA. On the theory of inconsistent formal systems. **Notre dame journal of formal logic**, v. 15, n. 4, p. 497-510, 1974.

GÖDEL, Kurt. Zum intuitionistischen aussagenkalkül. **Anzeiger Akademie der Wissenschaften Wien, mathematisch-naturwissenschaftliche Klasse**, v. 69, p. 65-66, 1932.

HJORTLAND, Ole Thomassen. Logical Pluralism, Meaning-Variance and Verbal Disputes. **Australasian Journal of Philosophy**, v. 91, n. 2, p. 355-373, 2013.

KNEALE, William; KNEALE, Martha. **The development of logic**. Oxford University Press, 1962.

ŁOŚ, Jerzy; SUSZKO, Roman. Remarks on sentential logics. **Indagationes mathematicae**, v. 20, n. 2, p. 177-183, 1958.

MALINOWSKI, Grzegorz. Inferential many-valuedness. In: **Philosophical logic in Poland**. Springer, Dordrecht, 1994. p. 75-84.

MALINOWSKI, Grzegorz. Many-valued logic and its philosophy. **The Many-Valued and Nonmonotonic Turn in Logic**, v. 8, p. 13-94, 2007.

MALINOWSKI, Grzegorz. **Many-Valued Logics**. Oxford Logic Guides 25. 1993.

MALINOWSKI, Grzegorz. Non-Fregean Logic and Other Formalizations of Propositional Identity'. **Bulletin of the Section of Logic**, v. 14, n. 1, p. 21-27, 1985.

MALINOWSKI, Grzegorz. Q-consequence operation. **Reports on Mathematical** Logic, v. 24, p. 49 – 59, 1990.

MARCOS, João. What is a non-truth-functional logic?. **Studia Logica**, v. 92, n. 2, p. 215, 2009.

PIRÓG-RZEPECKA, Krystyna. **Systemy nonsense-logics**. Państowe Wydawn. Naukowe, 1977.

PRIEST, Graham. The logic of paradox. **Journal of Philosophical logic**, v. 8, p. 219-241, 1979.

PRIEST, Graham. **In contradiction**. Oxford University Press, 2006a.

PRIEST, Graham. **Doubt truth to be a liar**. Oxford University Press, 2006b.

PRIEST, Graham. **An introduction to non-classical logic: From if to is**. Cambridge University Press, 2008.

QUINE, Willard V. **Philosophy of logic**. Harvard University Press, 1986.

RESCHER, Nicholas. **Many-valued logics**. McGraw Hill New York, 1969.

SMITH, Peter. Squeezing arguments. **Analysis**, v. 71, n. 1, p. 22-30, 2011.

SUSZKO, Roman. The Fregean axiom and Polish mathematical logic in the 1920s. **Studia Logica: An International Journal for Symbolic Logic**, v. 36, n. 4, p. 377-380, 1977.

TARSKI, Alfred. **Logic, semantics, metamathematics: papers from 1923 to 1938**. Hackett Publishing, 1983.

TSUJI, Marcelo. Many-valued logics and Suszko's thesis revisited. **Studia Logica**, v. 60, n. 2, p. 299-309, 1998.

WÓJCICKI, Ryszard. **Theory of logical calculi: basic theory of consequence operations**. Springer Science & Business Media, 2013.

WITTGENSTEIN, Ludwig. **Tractatus Logico-Philosophicus.** Tradução de Luiz Henrique Lopes dos Santos. São Paulo, Edusp, 2010.

PLURALISMO E INFERENCIALISMO: UMA APOLOGIA

Hermógenes Oliveira[1]

RESUMO

Tanto o pluralismo quanto o inferencialismo podem ser motivados por meio de contraposição à visão tradicional. Seja expandindo o inferencialismo para além das suas origens monistas, seja aderindo à uma abordagem mais pragmatista do pluralismo, obtemos uma combinação de ambas correntes que resulta numa postura coerente, robusta e proveitosa.

Palavras-chave: Inferencialismo. Pluralismo. Lógicas filosóficas. Paradoxos da implicação material.

ABSTRACT

Both pluralism and inferentialism can be motivated by objections to the traditional view. Either by extrapolating inferentialism from its monist beginnings, or subscribing to a more pragmatist stance on pluralism, we obtain a combination of both views that is arguably promising, coherent and robust.

Keywords: Inferentialism. Pluralism. Philosophical logics. Paradoxes of material implication.

1. Pluralismo

Já nos primórdios da lógica moderna, na virada do século XX, notava-se certa inadequação daquilo que hoje se conhece por lógica clássica com relação ao uso das respectivas expressões lógicas na linguagem natural. Russell (1903, §37), por exemplo, escreveu:

> Mas, ao desenvolver as consequências dos nossos pressupostos sobre a implicação, nós fomos levados à conclusões que, de maneira alguma, concordam com aquilo que é comumente aceito com respeito à implicação, pois nós concluímos que qualquer proposição falsa implica toda proposição e qualquer proposição verdadeira é implicada por toda proposição.

Os problemas de relevância da implicação clássica são bem conhecidos e amplamente discutidos na literatura na forma de "paradoxos da implica-

[1] Professor da Universidade Federal da Paraíba.
E-mail: hermogenes@cchla.ufpb.br. ORCID: 0000-0002-9251-6973.

ção material" (SANZ, 2000). A defesa desse tipo de inadequação da lógica clássica, especialmente no contexto dos programas de fundamentação da matemática, frequentemente apelava a questões de conveniência ou praticidade formal, sentimento muito bem expresso por Gentzen (1936, §5.24, ênfases no original) ao discutir a regra estrutural de enfraquecimento no cálculo de sequentes:

> Uma modificação estrutural conforme 5.243 significa que se pode *adicionar a uma asserção uma hipótese qualquer*, da qual a asserção passa também a depender junto com as outras eventuais hipóteses. À primeira vista, isso parece estranho. Porém, é inevitável conceder que, caso uma asserção seja *correta*, ela permanece válida sob o pressuposto de uma *hipótese qualquer*. (Caso se queira exigir que isso só valha para os casos de uma "efetiva dependência", surgiriam então consideráveis dificuldades por causa de provas nas quais hipóteses tenham uso apenas *aparente*)

Embora considerações de praticidade formal sejam até legítimas no contexto matemático do surgimento da lógica moderna, o balanço já não se mostra tão favorável no contexto da lógica entendida como teoria da argumentação ou do raciocínio. Neste último contexto, o âmbito teórico é bem mais amplo e há uma ênfase maior em adequação conceitual e capacidade explicativa do que em simplicidade formal e robustez técnica.

A insatisfação com aspectos variados da lógica clássica motivou o surgimento de uma pluralidade das chamadas *lógicas filosóficas* (HAACK, 2002). Algumas dessas lógicas se baseiam no questionamento de princípios clássicos (lógica intuicionista, lógica relevante), outras em limitações da lógica clássica em capturar certos tipos de inferências (lógicas modais), outras ainda se baseiam em ambas as motivações anteriores (lógica modal intuicionista).

Em face desse cenário, existe uma tendência em se considerar a pluralidade de lógicas filosóficas não só como rivais da lógica clássica, mas também como rivais entre si. Isso advém da intuição de que as regras da argumentação ou do raciocínio são únicas e, portanto, apenas uma proposta pode estar correta. Contudo, há duas questões problemáticas na sustentação dessa visão:

1. Numa certa medida, a intuição de que apenas uma lógica pode estar correta se apoia na ideia de que as diferentes lógicas são irremediavelmente incompatíveis. Mas isso não é bem assim. Muitas lógicas distin-

tas podem ser traduzidas e reinterpretadas umas nas outras, abrindo espaço para alguma espécie de compatibilização entre elas.

2. A universalidade e a neutralidade com relação ao tópico também são comumente aventadas em favor de um monismo lógico. Ainda que eventualmente compatibilizadas, assim reza o argumento, o resultado seria um amálgama de lógicas especializadas, cada uma dedicada ao seu nicho: lógica clássica para a análise matemática clássica, lógica intuicionista para computabilidade, lógica epistêmica para o raciocínio sobre agentes e suas crenças, e assim por diante. Cabe lembrar, no entanto, que nem a universalidade nem a neutralidade podem ser consideradas como irrestritas, sob pena de infactibilidade ou trivialidade. Nesse sentido, dentre outras questões, alguns princípios fundamentais da conjunção e da disjunção no âmbito afirmativo não se sustentam em raciocínios envolvendo imperativos e perguntas. A regra de introdução da disjunção, por exemplo, não se revela apropriada quando a disjunção conecta orações imperativas. Porém, se no afã de alcançar universalidade absoluta, nos permitirmos esse tipo de contraexemplo, a empreitada será inevitavelmente malfadada.

Consideremos ambos os pontos acima um pouco mais detidamente. O primeiro ponto foi bastante arguido por Quine (1972), com a sua concepção de que mudar de lógica é mudar de assunto. É curioso, porém, que um anti-excepcionalista lógico como Quine, que acreditava que a lógica fosse contínua com a ciência, subscrevesse a essa concepção, pois certamente rejeitaria a ideia de que teorias científicas concorrentes, digamos, sobre a natureza do elétron, tratam de assuntos diferentes.

Bem, anti-excepcionalismo à parte, seria possível argumentar que a suposta centralidade da lógica para as empreitadas do raciocínio, argumentação e justificação faz com que um embate entre duas lógicas genuinamente distintas resulte infalivelmente em discussão cruzada. Entretanto, nenhuma única lógica disponível no mercado atualmente parece fazer jus a um papel tão central. Afinal, as diferenças, vantagens e desvantagens entre as mais vari-

adas lógicas podem ser facilmente discutidas sem que ocorra qualquer falha de comunicação, o que indica que elas ao menos não se limitam a capturar certo núcleo lógico absolutamente indispensável.

Caminhando então para o segundo ponto, que tipo de lógica seria essa que desfruta de universalidade e neutralidade suficientes para ocupar um lugar tão central, e potencialmente normativo, para as práticas do raciocínio e da argumentação? Pergunta difícil, certamente. Principalmente se almejamos certa adequação satisfatória com a prática em vez de regimentá-la para os propósitos de um programa específico (e.g. a fundamentação da matemática). O problema é que nossas práticas argumentativas são tão versáteis, variadas e múltiplas que um monista, por mais reformador que seja, pode acabar se revelando demasiado dogmático.

Nesse sentido, destacam-se pelo menos duas tendências. Por um lado, podemos nos deixar conduzir rumo à universalidade e neutralidade. O movimento natural, neste caso, é de retração a um núcleo indispensável, na medida em que concedemos limitações e particularidades no uso de princípios lógicos. Trata-se de uma tendência *minimalista*. Por outro lado, podemos nos deixar conduzir rumo à ampliação e conciliação. A experiência indica que esta abordagem precisa conviver com alguma medida de especialização e pluralidade, pois ocorre que certas lógicas precisam estar calibradas aos seus contextos e, outras vezes, elas podem não ser passíveis de conciliação completa com as demais. É uma troca razoável, caso estejamos interessados em levar a lógica para os mais variados cenários de aplicação do raciocínio. Trata-se, portanto, de uma tendência *expansionista*.

Neste ensaio, o pluralismo lógico consiste numa posição pacífica frente a uma pluralidade de lógicas. Essa posição pode decorrer da tese de que existem várias lógicas igualmente corretas (RUSSELL, 2019), mas também de uma postura mais niilista que não considera as diversas lógicas como concorrentes e/ou não considera suficientemente objetiva uma disputa entre elas por um único posto de correção cabal. Independente da razão subjacente, o pluralismo entrosa perfeitamente com uma tendência expansionista ao mesmo tempo em que não obstrui um programa minimalista. Inclusive, o pluralismo

se casa tão bem com um programa expansionista que o interesse de alguém num programa desses pode, eventualmente, ser a principal motivação para uma adesão ao pluralismo, ainda que a contragosto. Já um programa minimalista, na medida em que busca traçar conexões, traduções e compatibilizações entre lógicas, não é necessariamente adverso a uma postura pluralista. O pluralismo é compatível com o projeto de confecção de um mapa de conexões entre os princípios e sistemas lógicos no qual, possivelmente, um núcleo essencial se revele.

Oposto ao pluralismo, nós temos o monismo, tanto o de vertente conservadora quanto o de vertente reformadora. Frequentemente, o monista defende não apenas que há uma única lógica correta, mas já indica também qual seria ela. O conservador atribui o título à lógica clássica, possivelmente com algumas concessões à lógicas complementares dela (e.g. algumas lógicas modais). Por outro lado, o reformador atribui o título a uma lógica filosófica (intuicionista, relevante, intuicionista relevante).

Este ensaio começou com algumas observações sobre a inadequação da lógica clássica no que concerne a prática. Como observado, essas limitações se justificam no contexto de certos programas investigativos. No contexto de uma teoria geral da argumentação, não há porque permitir à lógica clássica qualquer privilégio, ressalvadas as vantagens acumuladas, principalmente em termos de maturidade e amplitude teórica, devido a sua prevalência nos últimos séculos. Neste contexto, talvez os fatores mais influentes no monismo conservador sejam mesmo a mera tradição e o dogmatismo.

Também o reformador, *quando monista*, está sujeito a objeções. A reforma, de fato, pode trazer um maior equilíbrio para o balanço entre a teoria e a prática, e, nesse sentido, configurar um ganho em relação ao monismo conservador. Porém, a tensão entre a tendência minimalista, subjacente à reforma, e a necessidade de cobrir um corpo palpável de práticas inferenciais, conduz a trocas inevitáveis e, muitas vezes, infelizes. Um exemplo paradigmático é a situação num certo tipo de lógica relevante, onde é necessário se escolher entre o silogismo disjuntivo e a regra do corte, ambas as quais se gostaria de reter (TENNANT, 2015, §1.2). Em contraste, o pluralista, por não estar compelido

a uma escolha única *absoluta*, consegue conviver com tais infortúnios de maneira mais graciosa.

Uma questão de demarcação está implícita no contraste entre pluralismo e monismo. O que é, afinal, uma expressão *lógica*? E qual é a diferença entre uma expressão lógica e as demais? O que significa para uma inferência ter força *dedutiva* ou lógica? Qual são as fronteiras da lógica? Lógica modal alética é lógica pura, ou inclui aspectos de metafísica dos mundos possíveis? Lógica epistêmica é lógica pura, ou inclui aspectos de epistemologia?

O monista está mais estreitamente comprometido com a questão da demarcação, pois relações inferenciais que não estejam capturadas pela lógica correta estão automaticamente relegadas ao campo extra-lógico. Já o pluralista, dependendo dos detalhes da sua confissão, pode oportunamente traçar algumas divisões, mas dispõe de maior flexibilidade por não estar atado de antemão. Além do mais, um pluralista anti-excepcionalista, por não conferir status especial à lógica, pode inclusive se esquivar do problema de demarcação completamente.

Enfim, trocando em miúdos, o pluralismo se mostra uma postura mais razoável e vantajosa do que o monismo. Muitas das considerações e argumentos dos monistas, sobretudo dos monistas reformadores, no que concerne as virtudes e superioridade das suas respectivas lógicas, continuam eficazes para o pluralista, pois ele não está constrangido ao relativismo e pode muito bem apreciar comparações entre as alternativas. Ademais, o pluralista está menos ameaçado por problemas notadamente difíceis, como os da demarcação, incompatibilidade de princípios intuitivamente adequados e a versatilidade e sensibilidade contextual do raciocínio prático. Contudo, mais que uma simples artimanha para se evitar questões espinhosas, o pluralismo pode ser considerado uma espécie de herdeiro do conhecido *princípio de tolerância* (CARNAP, 1934, §17) e, nesse sentido, representa uma vitória da diplomacia sobre o dogmatismo.

2. Inferencialismo

Na lógica proposicional clássica, o significado dos conectivos é explicado por meio de uma análise combinatória dos valores semânticos possíveis: verdadeiro ou falso. Cada conectivo então representa uma certa função desses valores, algo que é exibido de maneira bastante perspicaz nas conhecidas tabelas. Essa análise combinatória é matematicamente interessante e desfruta de certas aplicações práticas (por exemplo, no desenho de chaveamentos e circuitos elétricos).

No entanto, a posição dominante que essa combinatória clássica habitualmente desfruta no âmbito da análise do raciocínio, sendo inclusive ensinada em introduções à lógica para alunos dos cursos de filosofia, baseia-se, em grande parte, em uma certa concepção da linguagem e seus respectivos princípios semânticos. Trata-se de uma concepção figurativa (ou representativa) na qual nomes se relacionam com objetos, predicados com propriedades e proposições com valores aléticos (SÀÁGUA, 2005).

Tradicionalmente, essa concepção está mergulhada ainda num arcabouço composto de várias teses semânticas, como a bivalência (toda proposição ou bem é verdadeira ou é falsa), a tese da existência e univocidade da contraditória (a toda proposição significativa corresponde uma única proposição contraditória, que também é significativa) e o princípio de não-contradição (uma proposição e sua contraditória não se verificam simultaneamente).

Sustentada por esse pano de fundo, a lógica proposicional clássica é frequentemente ensinada e aplicada em conexão com a prática da argumentação e do raciocínio em geral, embora sua vocação para esse âmbito seja bastante tênue. Portanto, é compreensível que se questione a legitimidade do ensino de combinatória booleana para alunos ingressantes dos cursos de filosofia, por exemplo, pois, sem uma fundação sólida no vasto cenário teórico indicado acima, a combinatória é, realmente, tudo o que se tem.

Quem se ocupa de lecionar disciplinas de introdução à lógica em cursos de filosofia está ciente das dificuldades do alunado, especialmente com a implicação clássica. A correspondência entre a implicação clássica e o uso das respectivas expressões ("se… então…", "implica" e etc.) é bastante frágil e,

nas linhas da tabela onde a proposição antecedente é falsa, até mesmo contraintuitiva (WASON, 1968). Uma estratégia frequentemente adotada é apoiar-se na conjunção e na negação, que são intuitivamente mais robustas, para se apresentar a disjunção e a implicação, sempre se apoiando na análise combinatória da tabela na discussão de exemplos muito contraintuitivos. Também é possível refinar o contexto dos exemplos de modo a induzir uma interpretação mais clássica (EVANS, 2002).

Lamentavelmente, contudo, não é incomum que o desencaixe entre a prática do raciocínio e a lógica clássica seja tomada, não como evidência da inadequação desta última, mas como base para especulações e ilações descuidadas sobre a inteligência ou racionalidade do ser humano ou de um grupo de pessoas (WILKINS, 1928).

Ora, convém ressaltar que a intenção deste ensaio não é declarar a supremacia da prática sobre a teoria ou a falência da idealização e regimentação teórica. Dito isso, cabe perfeitamente questionar até que ponto a perpetuação do paradigma clássico não está mais a serviço da tradição e do dogma do que da virtude teórica, propriamente dita. Vale lembrar ainda que parte do apelo inicial de abordagens tradicionais se deve ao fato de que sua transmissão inquestionada na formação das novas gerações frequentemente obscurece desproporcionalmente as alternativas concorrentes. Assim, é natural que um ônus exagerado recaia sobre abordagens desafiantes.

No que concerne a lógica como teoria da argumentação e do raciocínio, o inferencialismo ocupa uma posição de destaque, pois se sensibiliza, desde o início, com o *uso* das expressões. Em contraste com a concepção figurativa tradicional, o inferencialismo se inspira no chavão wittgensteiniano que relaciona *significado* e *uso*. Essa concepção semântica é notadamente mais pragmática, e tem sido explorada por diversos autores (BRANDON, 2013; PEREGRIN, 2014). No âmbito da lógica propriamente dita (SCHROEDER-HEISTER, 2018), ela normalmente ganha contornos particulares mediante aportes teóricos advindos da chamada *Teoria da Demonstração*, em especial do trabalho de Gentzen (1935).

Assim como a abordagem clássica tradicional se apoia em princípios semânticos fundamentais como a bivalência, o inferencialismo lógico também exibe princípios reguladores como a noção de *harmonia dedutiva* ou *harmonia inferencial* (RUBIN, 2017). No entanto, o arcabouço justificatório é completamente distinto, pois o critério de harmonia se impõe por considerações de consistência pragmática e razoabilidade das práticas inferenciais.

Nós não entraremos nos detalhes da proposta inferencialista. Não obstante, como caracterização rudimentar para os propósitos estreitos deste ensaio, trata-se de uma abordagem semântica na qual o significado da disjunção, por exemplo, é dado por um emprego inferencial emblemático, como o raciocínio por casos, em vez do seu comportamento com relação à veracidade das proposições envolvidas, como ocorre na abordagem clássica. Notamos já algumas vantagens. O raciocínio por casos é bastante intuitivo e, inclusive, corriqueiramente utilizado na forma de silogismos disjuntivos. Além do mais, o raciocínio por casos é compatível com interpretações inclusivas e exclusivas da disjunção (MILNE, 2012, §2), enquanto uma tabela clássica completa para a disjunção deve optar por uma delas.

Agora, com relação à disjunção, qual parece mais robusto e esclarecedor no contexto de uma teoria da argumentação: o raciocínio por casos ou uma tabela combinatória que sequer se aplica muito bem em alguns casos? É uma pergunta retórica, claro.

Tomar o raciocínio por casos como determinante do significado da disjunção pode causar uma certa estranheza inicial, pois a abordagem tradicional, na qual todos fomos treinados, costuma fazer o caminho inverso: a semântica da tabela nos fornece o significado e, com base nele, nós justificamos o raciocínio por casos. Contudo, com a licença de mais algumas perguntas retóricas, a ordem de explicação não deveria ser a inversa? Se a razoabilidade da tabela da disjunção se estabelece por sua capacidade de justificar o raciocínio por casos, a despeito de inadequação em alguns outros contextos, por que não descartar a tabela e tomar diretamente o raciocínio por casos? Uma resposta comum é: "Porque o raciocínio por casos é uma regra de inferência e regras de inferência fazem parte da sintaxe e não da semântica". Bem, se o

pressuposto é que somente a abordagem tradicional é digna do título de semântica, então discussão encerrada. Com a mente um pouco mais aberta, no entanto, poderíamos ainda nos questionar: Mas como distinguiríamos então um sistema formal da sua semântica? Como demonstraríamos a adequação de um em relação ao outro? É possível, em semânticas inferencialistas, obter resultados técnicos como completude e correção?

Todas essas questões são motivadas por uma associação equivocada entre o inferencialismo lógico e o mecanismo sintático. Do ponto de vista de um sistema sintático formal, o raciocínio por casos, sozinho, não dá conta do fragmento disjuntivo de nenhuma lógica. Portanto, a adequação de um sistema formal a uma certa semântica inferencialista, que pode ou não se valer de inferências emblemáticas retiradas do sistema formal, é uma questão bem formulada e pode, sim, receber tratamento e solução técnica por meio de teoremas de completude e correção (PIECHA, 2016). Portanto, os lógicos de viés mais matemático nada tem a temer do inferencialismo.

A ostentação de uma suposta concordância estreita com a prática, conforme pode se ler acima, também pode conduzir à objeção de que semânticas inferencialistas não oferecem qualquer ganho teórico, uma vez que estão submetidas às vicissitudes e inconsistências da prática, sem qualquer pretensão de regimentá-la. Em geral, esta também é uma caracterização injusta, embora um balaço saudável entre regimentação e perspicácia pragmática seja realmente um objetivo almejado, ainda que elusivo.

No âmbito de algumas lógicas filosóficas, o inferencialismo lógico propicia maior sofisticação conceitual, o que pode levar inclusive a ganhos técnicos. As propostas inferencialistas pioneiras, por exemplo, surgiram num cenário monista reformador calibrado para a lógica intuicionista. Antes do seu advento, as principais abordagens semânticas para a lógica intuicionista eram a interpretação BHK e a semântica de Kripke (MOSCHOVAKIS, 2018). A interpretação BHK resiste tratamentos técnicos por causa do seu caráter intrinsecamente informal. A semântica de Kripke, por outro lado, apelava a uma metafísica de estágios possíveis de conhecimento, um assunto aparentemente ortogonal ao da validade intuicionista. Ademais, o resultado de comple-

tude na semântica de Kripke apela para raciocínios estritamente clássicos. Em contraste, as principais propostas inferencialistas para a lógica intuicionista se baseiam apenas em regras inferenciais bastante incontroversas, oferecendo assim uma abordagem semântica conceitualmente mais enxuta e construtivamente mais agradável.

A despeito das suas origens, porém, a abordagem inferencialista está longe de restringir-se à lógica intuicionista. Propostas variadas indicam como se poderia obter harmonia dedutiva no contexto clássico (READ, 2000). Resultados analíticos em dedução natural e cálculo de sequentes para diversas lógicas filosóficas, modais e subestruturais, também abrem a porta para tratamentos inferencialistas correspondentes (DOŠEN, 1989).

3. Pluralismo inferencialista (ou inferencialismo pluralista)

Existem duas rotas. Ambas conduzem a um mesmo programa, objeto apologético deste ensaio. Uma rota adentra o pluralismo e dali caminha rumo ao inferencialismo. Outra rota parte das terras inferencialistas em direção a uma perspectiva mais plural.

O pluralismo inferencialista (ou inferencialismo pluralista, conforme o ponto de vista) é caracterizado por uma postura pluralista no tratamento com as diversas lógicas, conjugada com uma abordagem inferencialista na semântica. Essa parceria apresenta virtudes que não são alcançadas separadamente.

Em primeiro lugar, cabe ressaltar que o arcabouço gentzeniano composto por harmonia inferencial e os resultados analíticos correspondentes (normalização e eliminação do corte) de maneira alguma prescinde de uma rica variedade de fenômenos e sistemas lógicos. Nesse sentido, conforme observado acima, o monismo reformador dos pioneiros intuicionistas se contrasta com trabalhos sobre harmonia na lógica clássica e investigações subestruturais de sistemas gentzenianos. Esses desdobramentos se revelam genuinamente inferencialistas na medida em que mantêm uma adesão aos pilares essenciais: sensibilidade ao emprego inferencial das constantes lógicas, destilamento do significado por meio de regras emblemáticas do uso e respeito ao critério de harmonia.

Há que se conceder que a harmonia impõe um certo critério de demarcação, mitigando assim o escopo do pluralismo. Entretanto, especialmente para aqueles que reclamam de frouxidão regimental nas abordagens sensibilizadas pela prática, isso dificilmente tornará o inferencialismo pluralista mais antipático. Muito pelo contrário.

Partindo de uma perspectiva pluralista, a parceria também pode se mostrar proveitosa. As abordagens semânticas usuais de boa parte das lógicas filosóficas se valem da receita tradicional inaugurada por Tarski (2009) e posteriormente expandida por Kripke (1963). A insistência nessa receita tradicional, cuja metateoria é invariavelmente clássica, resulta, em alguns casos, em abordagens ditas "semânticas" de complexidade assustadoramente elevada para lógicas que, supostamente, buscam capturar as regras do raciocínio na sua roupagem mais prática e intuitiva.

Talvez um dos casos mais paradigmáticos nesse sentido seja o das lógicas relevantes. A definição de validade na semântica considerada tradicional, desenvolvida por Routley e Meyer a partir da semântica de reticulados de Urquhart (Sua cabeça já está doendo? Continue lendo que piora), começa com algo do tipo "Uma moldura F de Routley-Meyer para uma linguagem proposicional é um quádruplo (W,R,*,0), onde W é um conjunto não vazio, R é uma relação ternária sobre W,…" e só vai piorando quanto mais adiante se segue. De fato, o contraste entre o forte apelo intuitivo das motivações iniciais e o surpreendente embaraço matemático do respectivo desenvolvimento formal dificilmente passa despercebido aos leitores dos tomos clássicos de Anderson e Belnap (1975, 1992). Por outro lado, para capturar a implicação relevante, um inferencialista necessita apenas revogar o descarte vácuo de hipóteses, numa conexão clara e direta com a motivação intuitiva subjacente (PRAWITZ, 1965, §VII.1).

Aliás, o potencial da abordagem inferencialista não se limita ao enxugamento do aparato matemático. Há ainda uma promessa de economia no âmbito conceitual. Conforme já mencionado com relação ao caso intuicionista, o inferencialista não precisa se ocupar de questões envolvendo a natureza ou o status metafísico das entidades habitantes do conjunto não vazio W.

Não obstante, apesar dos prospectos bastante promissores, as investigações na intersecção do pluralismo e do inferencialismo ainda se encontram em estágio incipiente. Algumas das virtudes aludidas acima se apoiam em rascunhos programáticos cujos detalhes ainda não se materializaram. Nesse sentido, este curto ensaio apologético pretende ser, sobretudo, uma chamada ao trabalho.

Referências

ANDERSON, A. R.; BELNAP, N. *Entailment: The Logic of Relevance and Necessity*. Princeton: PUP, 1975 (Vol. 1), 1992 (Vol. 2).

BRANDON, R. *Articulando Razões: Uma Introdução ao Inferencialismo*. Porto Alegre: Editora PUCRS, 2013.

CARNAP, R. *Logische Syntax der Sprache*. Wien: Springer, 1934.

DOŠEN, K. "Logical Constants as Punctuation Marks". *Notre Dame Journal of Formal Logic*, vol. 30, n. 3, 1989, p. 362-381.

EVANS, J. St. B. T. "Logic and human reasoning: an assessment of the deduction paradigm". *Psychological Bulletin*, vol. 128, n. 6, 2002, p. 978-996.

HAACK, S. *Filosofia das Lógicas*. São Paulo: Editora UNESP, 2002.

KRIPKE, S. "Semantical analysis of modal logic I". *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 9, 1963, p. 67-96.

GENTZEN, G. "Die Widerspruchsfreiheit in der reinen Zahlentheorie". *Mathematische Annalen*, vol. 112, 1936, p. 493-565.

GENTZEN, G. "Untersuchungen über das logische Schlißen I". *Mathematische Zeitschrift*, vol. 39, n. 2, 1935, p. 156-210.

MILNE, P. "Inferring, splicing and the Stoic analysis of argument". In NOVAES, C. D.; HJORTLAND, O. T. (eds.) *Insolubles and Consequences: Essays in Honour of Stephen Read*, Londes: College Publications, 2012.

MOSCHOVAKIS, J. "Intuitionistic logic". In ZALTA, E. N. (ed.) *Stanford Encyclopedia of Philosophy*, MRL: Stanford University, 2018. Disponível em: https://plato.stanford.edu/entries/logic-intuitionistic/. Acesso em: 30 de maio de 2020.

PRAWITZ, D. *Natural Deduction: A Proof-Theoretical Study*, Estocolmo: Almqvist & Wiksell, 1965.

PEREGRIN, J. *Inferentialism: Why Rules Matter*. Berlin: Springer, 2014.

PIECHA, T. "Completeness in proof-theoretic semantics". In PIECHA, T.; SCHROEDER-HEISTER, P. (eds.) *Advances in Proof-Theoretic Semantics*, Berlin:Springer, 2016. Disponível em: https://link.springer.com/978-3-319-22686-6. Acesso em: 30 de maio de 2020.

QUINE, W. O. *Filosofia da Lógica*. Rio de Janeiro: Zahar Editores, 1972.

READ, S. "Harmony and autonomy in classical logic". *Journal of Philosophical Logic*, vol. 29, 2000, p. 123-154.

RUBIN, M. "Constantes lógicas y la armonía de las reglas de inferencia". Revista de Humanidades de Valparaíso, vol. 5, n. 9, 2017, p. 103-119. Disponível em: https://revistas.uv.cl/index.php/RHV/article/view/64. Acesso em: 30 de maio de 2020.

RUSSELL, B. *The Principles of Mathematics*. Cambridge: CUP, 1903.

RUSSELL, G. "Logical pluralism". In ZALTA, E. N. (ed.) *Stanford Encyclopedia of Philosophy*, MRL: Stanford University, 2019. Disponível em: https://plato.stanford.edu/entries/logical-pluralism/. Acesso em: 30 de maio de 2020.

SÀÁGUA, J. "Lógica de primeira ordem". In BRANQUINHO, J.; MURCHO, D.; GOMES, N. G. *Enciclopédia de Termos Lógico-Filosóficos*, São Paulo: Martins Fontes, 2006.

SANZ, W. "Problemas do condicional: relevância". *Filósofos*, vol. 5, n. 1, 2000, p. 175-185.

SCHROEDER-HEISTER, P. (2018). "Proof-theoretic semantics". In ZALTA, E. N. (ed.) *Stanford Encyclopedia of Philosophy*, MRL: Stanford University, 2018. Disponível em: https://plato.stanford.edu/entries/proof-theoretic-semantics/. Acesso em: 30 de maio de 2020.

TARSKI, A. *A Concepção Semântica da Verdade*. São Paulo: Editora UNESP, 2009.

TENNANT, N. "Cut for Classical Core Logic". *The Review of Symbolic Logic*, vol. 8, n. 2, 2015, p. 236-256.

WASON, P. "Reasoning about a rule". *Quarterly Journal of Experimental Psychology*, vol. 20, n. 3, 1968, p. 273-281.

WILKINS, M. C. "The effect of changed material on the ability to do formal syllogistic reasoning". *Archives of Psychology*, vol. 102, 1928.

A QUESTÃO DA IDENTIDADE DE PROVAS, A TESE DA NORMALIZAÇÃO E ALGUNS DE SEUS PRESSUPOSTOS SEMÂNTICOS BÁSICOS

Tiago de Castro Alves[1]

RESUMO

O artigo oferece , em primeiro lugar, uma caracterização da questão da identidade de provas como um problema concernente à semântica de argumentos dedutivos, pertinente à lógica e à filosofia da linguagem. Em seguida, é dada uma descrição dos principais aspectos formais e conceituais daquela que é provavelmente a proposta de abordagem formal dessa questão mais bem desenvolvida na literatura: a chamada *tese da normalização*, proposta em caráter conjectural em Prawitz 1971. Finalmente, alguns dos pressupostos semânticos que alicerçam a formulação dessa tese – componentes importantes da concepção *representacionalista* de provas que ela instancia – são expostos e discutidos quanto a suas possíveis motivações e consequências relevantes. Para além dos objetivos atinentes ao seu teor, este tem a meta pedagógica de tornar acessível e fomentar a consideração da questão da identidade de provas no seio da comunidade filosófica de língua portuguesa.

Palavras-chave: Identidade de provas. Semântica. Argumentos dedutivos. Derivações. Normalização.

ABSTRACT

The article offers, first, a characterisation of the question of identity of proofs as a problem concerning the semantics of deductive arguments, pertaining to logic and philosophy of language. In the sequel, a description of the main formal and conceptual aspects of the *normalisation thesis* – probably the most well developed formal approach to identity of proofs in the literature so far, proposed as a conjecture by Prawitz in 1971 – is given. Finally, some of the semantical assumptions that form the basis of the thesis' formulation – important components of the *representationalist* view of proofs which it instantiates – are exposed and discussed as to its relevant consequences and possible motivations. Apart from the goals directly related to its contents, the article has the pedagogical aim of facilitating and stimulating the consideration

[1] Professor de Filosofia do Departamento de Filosofia da UFMT.
E-mail: tiagorca@hotmail.com. ORCID: 0000-0002-5786-5561.

of the question of identity of proofs amongst the portuguese-speaking philosophical community.

Keywords: Identity of proofs. Semantics. Deductive arguments. Derivations. Normalisation.

Introdução: a semântica de argumentos dedutivos

Desde o advento da lógica de Port-Royal de Arnauld e Nicole, ainda no século XVII, e até a chamada "revolução" lógica do século XIX – ainda tão carente de uma historicização apropriada –, foi padrão a prática de dividir os tradicionais manuais de Lógica, então dedicados sobretudo à silogística aristotélica, em três partes ou "livros", cada um dedicado a um objeto diferente da tradicional "doutrina dos elementos"; respectivamente: dos conceitos; dos juízos; e, finalmente, das inferências. O famoso manual de Lógica de Kant, editado por G.B. Jäsche – traduzido na íntegra para o português[2] – é um exemplo. Esse fato é relativamente bem difundido entre estudantes de filosofia.

Também é relativamente comum que, ao longo de sua formação, estudantes de filosofia sejam no mínimo postos a par da existência de um volume vultoso de literatura – produzida sobretudo após a mencionada "revolução" por autores da filosofia da linguagem de orientação "analítica" –, dedicada ao tratamento da semântica daquilo que folcloricamente, por assim dizer, é compreendido como contraparte, no léxico contemporâneo, dos dois primeiros elementos mencionados da lógica tradicional dos modernos: primeiramente, nomes, termos, etc. – contrapartes contemporâneas dos *conceitos*; e posteriormente, proposições, sentenças, etc. – contrapartes contemporâneas dos *juízos*.

É precisamente o conjunto dessas observações que enseja estranheza na recepção do fato de que as contrapartes contemporâneas do terceiro elemento – derivações, deduções, etc., contrapartes das *inferências* – careçam quase que em absoluto de abordagens semânticas semelhantes da literatura filosófica contemporânea. Não obstante, essa escassez não é um fenômeno to-

[2] I. Kant, Lógica. Trad. por Guido Antônio de Almeida. (Série Estudos Alemães, 93). 2ª ed. Rio de Janeiro: Tempo Brasileiro, 1992.

talmente misterioso de um ponto de vista histórico-conceitual: suas razões são em larga medida explicáveis em termos dos propósitos e orientações filosóficas formalistas dos desenvolvimentos ligados à *Beweistheorie* do programa fundacional hilbertiano, em muito responsáveis pelos contornos metamatemáticos com que se desenvolveram e legaram as noções contemporâneas "paralelas" à da inferência moderna. O estudo do teor dessa história, contudo, não será objeto das presentes considerações. Mostraremos, antes, um caminho atualmente aberto para o tratamento atual e futuro desse que ainda é um *desideratum*[3] no catálogo temático da filosofia da linguagem contemporânea, a saber: a semântica dos argumentos dedutivos. Mais especificamente, o esforço a ser aqui empreendido será o de cumprir as seguintes tarefas: 1. apresentar a chamada questão da "identidade de provas" como um problema filosófico; 2. sob o pano de fundo dessa caracterização do problema, apresentar aquela que considero a sua abordagem formal mais bem desenvolvida até hoje na literatura: a chamada tese da normalização. 3. discutir aspectos notáveis conceituais e matemáticos dessa abordagem do problema, em especial: (a) alguns fatos formais habitualmente usados como argumentos a favor da tese; e (b) a concepção denotacionalista da semântica das derivações que sua formulação original pressupõe.

1. A questão da identidade de provas

Nosso primeiro passo neste esforço será caracterizar *a questão* da identidade de provas, a fim de possibilitar uma apreciação adequada de seu sentido e sua relevância como um problema filosófico. A formulação preliminar a ser instrumentalmente usada como primeiro passo rumo à clareza a respeito dessa questão é a seguinte: Dado que Π é uma prova e que Σ é também uma prova, quando é que dizemos que Π e Σ são *a mesma* prova?

[3] Diagnóstico sugerido pelo prof. Peter Schroeder-Heister.

1.1. O que se prova vs. como se prova

Uma indagação plausível e relevante que a formulação acima pode suscitar é: por que alguém se perguntaria algo assim?

Há alguns fatos cuja observação suscita ou, no mínimo, enseja esse exato tipo de questão. O prof. Luiz Carlos Pereira costuma pontuar que esse tipo de questão atinge diretamente e de maneira muito concreta a vida prática de pesquisadores e acadêmicos: de fato, decisões a respeito da identidade ou não entre provas oferecidas para um mesmo resultado matemático, por exemplo, são a base para a aceitação ou a rejeição de propostas de publicação de artigos em periódicos. Esse tipo de situação torna flagrante um fato que o condiciona, e cuja observação tem sido provavelmente a porta de entrada efetivamente mais utilizada para o interesse na questão da identidade de provas, a saber: é frequentemente possível provar (viz. oferecer argumentos válidos para) as mesmas coisas de maneiras significativamente diferentes entre si. Por exemplo: posso demonstrar que minha caneta não está na gaveta esquerda da escrivaninha abrindo a gaveta, inspecionando seu conteúdo exaustivamente e verificando que a caneta não é parte dele; ou também simplesmente verificando que a caneta está de fato sobre o tampo da escrivaninha. Isso permite concluir que, a fim de determinar se Π e Σ são a mesma prova, não é suficiente que consideremos apenas *o que* Π e Σ provam – o que é, em muitos contextos, suficientemente evidente –, mas também *como* o fazem. Isso, contudo, não significa que *qualquer* diferença entre duas maneiras de provar uma mesma coisa é significativa; o que evidencia que não é claro que espécie de diferenças entre duas maneiras de se provar uma mesma coisa são ou não relevantes para determinar se se trata de uma mesma prova ou de provas distintas.[4]

[4] Além disso, é também possível, em princípio, que provas oferecidas para coisas diferentes sejam a tal ponto análogas que a diferença entre aquilo que provam não embargue a identidade entre elas – pode-se facilmente conceber a ideia de um mesmo argumento formal válido aplicado a premissas distintas e levando a conclusões respectivamente distintas, por exemplo. Em outras palavras, é possível em princípio que a identidade entre o que é provado por Π e Σ não seja nem sequer condição necessária para a eventual identidade entre Π e Σ. Neste texto, no entanto, nos concentraremos em considerações sobre identidade de provas levadas a cabo a partir da pressuposição de que a identidade entre o que é provado por Π e Σ é condição necessária para a identidade entre Π e Σ.

Em suma: carecemos de clareza a respeito de critérios segundo os quais identificar ou diferenciar provas/argumentos válidos oferecidos para uma mesma coisa. Em consonância com a literatura que consideraremos na sequência, utilizaremos essa observação como motivação para a investigação do problema da identidade de provas neste esforço, a saber: a identidade entre provas de uma mesma coisa não é trivialmente garantida pelo mero fato de serem provas de uma mesma coisa. Isto é: a eventual identidade entre *o que é provado* por, respectivamente, Π e Σ, não é entendido como *condição suficiente* para a eventual identidade entre Π e Σ. A esse respeito, Kosta Došen 2003 coloca:

> Para que toda a área da teoria geral da prova [*general proof theory*] faça sentido, e em particular para a consideração da questão de critérios de identidade para provas, não devemos admitir que quaisquer duas derivações com as mesmas hipóteses e conclusão sejam equivalentes; isto é, não deve ser o caso que nunca haja mais que uma prova com dadas hipóteses e uma dada conclusão. Não fosse assim, nossa área seria trivial. Isso é o divisor de águas entre a teoria da prova e o resto da lógica, em que não se se ocupa de provas, mas, no máximo, de relações de consequência.[5] (DOŠEN, 2003, p.14)

Esse tipo de observação confere legitimidade e relevância à questão da identidade de provas na medida em que a torna irredutível à questão a respeito da identidade entre aquilo que as provas provam.

1.2. Provas: τί ἐστιν

Os mais variados tipos de coisas têm sido designados pelo rótulo "provas": provas de amor (ex.:o Taj Mahal), provas de inocência criminal (ex.: um álibi), provas de teoremas (ex.: a prova de Euclides da infinitude dos primos) – isso para não mencionar casos aparentemente mais distantes de nossos presentes interesses, como provas gastronômicas ou provas escolares trimestrais

[5] No original: "For the whole field of general proof theory to make sense, and in particular for considering the question of identity criteria for proofs, we should not have that any two derivations with the same assumptions and conclusion are equivalent, i.e. it should not be the case that there is never more than one proof with given assumptions and a given conclusion. Otherwise, our field would be trivial. This marks the watershed between proof theory and the rest of logic, where one is not concerned with proofs, but at most with consequence relations." Tradução livre do autor.

de matemática. Provas aqui parecem nos interessar *qua*, grosso modo, *razões suficientes para a admissão daquilo que provam.* Assim, por exemplo: o Taj Mahal é supostamente suficiente como razão para admitirmos o amor de Shah Jahan por Aryumand Banu Begam; um álibi é supostamente suficiente para admitirmos a inocência do réu, por exemplo; a argumentação de Euclides é supostamente suficiente para admitirmos a infinitude dos primos; etc.

Na literatura considerada neste esforço, o escopo de coisas compreendidas sob esse rótulo diminui consideravelmente – em geral, limita-se a consideração aos casos de provas que podem ser expressas, levadas a cabo ou representadas por meio de derivações formais. Não obstante, é precisamente no seio dessa literatura que discussões com vistas à delimitação mais precisa do assunto em questão fazem com que se multipliquem as categorias lógico-ontológicas sob as quais são entendidas as coisas que tratamos por provas: normas, atos, objetos, etc. Assim, parece impor-se à investigação da identidade de provas uma tarefa preliminar, a saber: determinarmos o que uma prova é. Somente assim, presumivelmente, será possível que tenhamos um parâmetro para julgar a adequação de um critério de identidade proposto – para diferentes coisas, diferentes critérios seriam ou não respectivamente cabíveis.

Por outro lado, é frequente nessa mesma literatura a manifestação da ideia de que uma resposta adequada à questão da identidade de provas é ela própria central para a compreensão adequada de o que uma prova é. Diante disso, parece sugerir-se a possibilidade de que estejamos condenados a perseguir a própria cauda nessa empreitada: pois, por um lado, a resposta à pergunta "o que é uma prova?" dependeria de uma resposta adequada à questão da identidade de provas; e, por outro, o julgamento da adequação de uma resposta à questão da identidade de provas dependeria de termos clareza prévia a respeito do que estamos tratando por provas. Como sair desse círculo?

Uma forma de dissolver – ou, no mínimo, contornar – esse aparente problema é atentar para o fato de que a clareza que temos que ter de antemão a respeito do que provas são a fim de julgar propostas de critérios de identidade de provas não é o mesmo tipo de clareza sobre provas cuja obtenção presumivelmente depende de uma boa resposta à questão da identidade de provas.

A fim de compreender melhor essa colocação, podemos primeiramente oferecer uma resposta adequada à seguinte questão, que talvez já tenha se insinuado: por que é que alguém julgaria que compreender o que uma prova é depende em algum sentido de uma resposta à questão da identidade de provas?

Uma breve explicação para isso pode ser dada com recurso ao *slogan* Quineano *"no entity without identity"*. Em um breve texto sobre o conceito de identidade[6], Sundholm chama a atenção para o seguinte:

> Considere os dois tipos $\mathbb{N}^+ \times \mathbb{N}^+$ de pares ordenados de inteiros positivos e $\mathbb{Q}^+$ de racionais positivos. Formalmente, eles têm o mesmo critério de aplicação:
>
> $$\frac{p:\mathbb{N}^+ \quad q:\mathbb{N}^+}{\langle p,q\rangle:\alpha}$$
>
> <2, 3> e <4, 6> são elementos iguais do tipo $\mathbb{Q}^+$, mas não do tipo $\mathbb{N}^+ \times \mathbb{N}^+$. A fim de individuar os tipos em questão, diferentes critérios de identidade são necessários: o tipo $\mathbb{N}^+ \times \mathbb{N}^+$ é individuado pelo critério de identidade
>
> $$\frac{p:\mathbb{N}^+ \quad q:\mathbb{N}^+ \quad r:\mathbb{N}^+ \quad s:\mathbb{N}^+ \qquad p=r:\mathbb{N}^+ \quad q=s:\mathbb{N}^+}{\langle p,q\rangle=\langle r,s\rangle:\mathbb{N}^+}$$
>
> e o tipo $\mathbb{Q}^+$ pelo critério
>
> $$\frac{p:\mathbb{N}^+ \quad q:\mathbb{N}^+ \quad r:\mathbb{N}^+ \quad s:\mathbb{N}^+ \qquad p\times s = q\times r:\mathbb{N}^+}{\langle p,q\rangle=\langle r,s\rangle:\mathbb{Q}^+}$$ [7]

---

[6] Sundholm, G.B. (1999) Identity: Propositional, criterial, Absolute, in The Logica 1998 Yearbook, Filosofia Publishers, Czech Academy of Science, Prague, pp. 20–26.

[7] No original: "Consider the two types $\mathbb{N}^+ \times \mathbb{N}^+$ of ordered pairs of positive integers and $\mathbb{Q}^+$ of positive rationals. Formally they have the same application criterion:

$$\frac{p:\mathbb{N}^+ \quad q:\mathbb{N}^+}{\langle p,q\rangle:\alpha}$$

<2, 3> and <4, 6> are equal elements of type $\mathbb{Q}^+$, but not of the type $\mathbb{N}^+ \times \mathbb{N}^+$. In order to individuate the types in question different criteria of identity are needed: the type $\mathbb{N}^+ \times \mathbb{N}^+$ is individuated by the identity criterion

$$\frac{p:\mathbb{N}^+ \quad q:\mathbb{N}^+ \quad r:\mathbb{N}^+ \quad s:\mathbb{N}^+ \qquad p=r:\mathbb{N}^+ \quad q=s:\mathbb{N}^+}{\langle p,q\rangle=\langle r,s\rangle:\mathbb{N}^+}$$

and the type $\mathbb{Q}^+$ by the criterion

(SUNDHOLM, 2009, p.20–21)

Uma observação análoga poderia ser aplicada ao presente contexto de discussão de provas: temos uma miríade de possíveis critérios de aplicação, dos quais, seguindo a já mencionada orientação da literatura, selecionamos, digamos, "levadas a cabo ou representadas por meio de derivações formais". Isso, evidentemente, responde *algo* a respeito de o que é uma prova; mas não tudo. De fato, a partir de uma resposta como essa – i.e., um critério de aplicação –, uma noção distinta e particular de prova (como um tipo) só pode ser dada pela associação a ela de um critério de identidade específico; para diferentes critérios de identidade, noções respectivamente diferentes de prova resultariam. Note-se que um critério de identidade é necessário não apenas para determinar a identidade das provas individuais "dentro" do tipo, mas também a identidade *do próprio tipo* em questão. É nesse sentido que Martin-Löf, segundo Sundholm, sugere a combinação das máximas de teoria dos tipos *"no entity without type"* e *"no type without identity"* como uma maneira fácil de reconstruir o *slogan* quineano *"no entity without identity"*. Pode-se entender as alegações de que uma resposta à questão da identidade de provas é essencial para a compreensão do que é uma prova no espírito desse tipo de reflexão. Por tabela, pode-se também compreender a partir disso uma saída para o círculo epistêmico caracterizado anteriormente: a clareza a respeito da noção de prova pressuposta para que se julgue a adequação de um critério de identidade não está condicionada pela aceitação prévia de um critério de identidade[8]; e a clareza a respeito de o que é uma prova obtida pela aceitação de um critério de identidade, ainda que condicionada pelo estabelecimento prévio de um critério de aplicação, é de natureza distinta, mais específica e informativa.

$$\frac{p:\mathbb{N}^+ \quad q:\mathbb{N}^+ \quad r:\mathbb{N}^+ \quad s:\mathbb{N}^+ \qquad p\times s = q\times r:\mathbb{N}^+}{\langle p,q\rangle = \langle r,s\rangle :\mathbb{Q}^+}$$ " Tradução livre do autor.

[8] Apesar de, como veremos adiante, essa clareza – i.e. o estabelecimento de um critério de aplicação – não ser suficiente como fundamento para uma escolha entre diferentes critérios de identidade a ele adequados.

A presente abordagem da questão da identidade de provas primará pela conexão desta com a tradição filosófica clássica que a precede e condiciona: entender provas (viz. argumentos válidos) como *expressões linguísticas* de forma adequada – na esteira, por exemplo, da compreensão da noção de *silogismo* oferecida por Aristóteles nos *Primeiros Analíticos* (A1, 24b18-22) – e, assim, tratar a questão da identidade de provas como concernente à equivalência semântica viz. *sinonímia* entre elas. Nesse sentido, o presente esforço entende a questão da identidade de provas, em consonância com as ideias transmitidas na introdução acima, como pertencente à *semântica* das expressões linguísticas em questão.[9]

2. A tese da normalização

Ainda que escassamente, de fato, algo de suficientemente profundo, abrangente e preciso *já* foi proposto na literatura contemporânea a respeito da questão da identidade de provas, e tem sido esporadicamente levado adiante desde, no mínimo, o início dos anos 70 do século XX. Essas contribuições, contudo, permanecem no olvido ao longo da formação da esmagadora maioria dos estudantes de filosofia – mesmo daqueles mais dedicados a assuntos ligados a lógica e filosofia da linguagem. Tentaremos aqui uma breve reconstrução daquela que julgo ser, por sua solidez filosófica, precisão e influência, o principal exemplo de abordagem de índole formal dessa questão: a chamada *tese da normalização*.

2.1. Alguns traços essenciais da tese da normalização

A formulação "oficial" da tese da normalização é proposta textualmente por Dag Prawitz em 1971 no artigo *Ideas and Results in Proof Theory* ("Ideias e Resultados em Teoria da Prova", sem tradução para o português) , pág.257: "*Two derivations represent the same proof if and only if*

[9] Em de Castro Alves, 2018, no capítulo I, é oferecida uma defesa da presença desse tipo de abordagem semântica num *locus* clássico da literatura filosófica canônica: o opúsculo *Die falsche Spitzfindigkeit der vier syllogistischen Figuren,* de Kant, traduzido para o português por Luciano Codato no volume *Escritos pré-críticos*, publicado pela editora Unesp em 2005.

they are equivalent"[10](PRAWITZ, 1971, p.257). A despeito da aparente trivialidade com que soa posta desta maneira, essa tese está, em muitos e importantes sentidos, longe de ser um truísmo.

Primeiramente, a noção de equivalência em termos da qual ela é formulada é bastante específica: ela concerne a derivações em dedução natural, e pode ser definida como o fechamento reflexivo, transitivo e simétrico da relação de redutibilidade imediata entre estas. As reduções consideradas são aquelas envolvidas na normalização de derivações; assim, uma derivação se reduz imediatamente a uma outra derivação (cf. PRAWITZ, 1971, Seção II.3.3) quando esta é obtida a partir daquela por meio da remoção de uma *fórmula máxima* (isto é, uma fórmula cujo conectivo principal é * que é a conclusão da aplicação de uma regra de introdução de * e premissa maior da aplicação de uma regra de eliminação de *). Por exemplo, no caso da conjunção, a derivação:

$$\dfrac{\dfrac{\begin{matrix}\Pi_1 & \Pi_2 \\ A & B\end{matrix}}{A \wedge B}}{\begin{matrix}A \\ \Pi_3\end{matrix}}$$

reduz-se imediatamente à derivação:

$$\begin{matrix}\Pi_1 \\ A \\ \Pi_3\end{matrix}$$

Derivações também se reduzem imediatamente por meio de *expansões imediatas*. Estas podem ser operadas sobre fórmulas mínimas de derivações normais e assim, conformá-las ao que Prawitz chama de *forma normal expandida,* em que todas as fórmulas mínimas são atômicas. Por exemplo: seja a seguinte uma derivação normal e $A \wedge B$ uma fórmula mínima nela:

[10] Em tradução livre: "Duas derivações representam a mesma prova se e somente se elas são equivalentes."

$$\begin{array}{c} \Pi_1 \\ A \wedge B \\ \Pi_2 \end{array}$$

Por meio de uma expansão imediata, ela se deixa reduzir[11] imediatamente à derivação abaixo[12]:

$$\dfrac{\dfrac{\begin{array}{c}\Pi_1 \\ A \wedge B\end{array}}{A} \quad \dfrac{\begin{array}{c}\Pi_1 \\ A \wedge B\end{array}}{B}}{\begin{array}{c} A \wedge B \\ \Pi_2 \end{array}}$$

Observe-se *en passant* que há uma certa analogia entre os dois tipos de redução recém-apresentados, a saber: ambas envolvem derivações em que, a partir de uma dada premissa, é inferida uma conclusão que não contribui, por assim dizer, para a obtenção da fórmula final a partir das iniciais. Em alguns casos, como o da conjunção, ocorre, de fato, um *retorno* imediato à referida premissa após a inferência da conclusão do tipo indicado. Assim, no primeiro caso, a derivação de A a partir de Γ a ser reduzida é caracterizada pela introdução e pela eliminação imediatamente subsequente de uma fórmula complexa, que é *removida* por meio da redução sem prejuízo da "boa-formação" da derivação A a partir de Γ. Dualmente, no segundo caso, a derivação reduzida é caracterizada pela eliminação e pela introdução imediatamente

[11] Pode causar alguma estranheza dizer que operar *expansões* imediatas sobre uma derivação a *reduz.* Em particular, uma expansão imediata não parece reduzir, mas sim *aumentar* a derivação, no sentido de que insere algo que não estava lá previamente. Em verdade, trata-se de um problema sedimentado na terminologia técnica: para o oposto de expansão, reserva-se o termo *contração*, e não *redução*. No jargão, este último se refere a operações de transformação cuja direção determina ou está em conformidade com a ordem de uma sequência ou cadeia de reduções. Por exemplo: a conversão de "3+2" em "5" é simultaneamente uma contração e uma redução quando se considera o cálculo da soma nos naturais; já a conversão de "18" em "2 x 3^2" é simultaneamente uma expansão e uma redução quando se considera o cálculo da fatoração prima dos naturais.

[12] Seria cabível perguntar-se por que a aplicação de expansões imediatas se restringe a esses casos. A razão, com efeito, não tem a ver com identidade de provas, mas, antes, com a *normalização* de derivações – que é, a propósito, a razão pela qual estas reduções foram concebidas. Ao fim e ao cabo, essa restrição sobre as expansões imediatas é precisamente o que permite a obtenção da normalização forte na presença das reduções por meio das quais fórmulas máximas são removidas (sobre essa questão e outras intimamente relacionadas, cf. C.-Jay, N.Ghani, *The virtues of eta-expansion.* J. Functional Programming 5 (2): 135-154, April 1995, Cambridge University Press).

subsequente de uma fórmula complexa, que surge a partir da *inserção* de uma ou mais fórmulas mais simples por meio da redução.[13]

Outras reduções por meio das quais uma derivação se deixa reduzir imediatamente são as chamadas *reduções permutativas*, que concernem às eliminações da disjunção e do quantificador existencial. Por meio delas, é possível remover os *segmentos máximos* (chamados "máximos" por uma razão similar à do caso das fórmulas máximas: trata-se de sequências de ocorrências repetidas de uma mesma dada fórmula, uma imediatamente sob a outra, tais que a primeira é a conclusão de uma aplicação de uma regra de introdução e a última é a premissa maior de uma aplicação de uma regra de eliminação. Note que fórmulas máximas podem ser definidas como um caso específico e limítrofe de segmentos máximos, nos quais apenas uma ocorrência da referida fórmula se dá. Segmentos máximos mais longos podem surgir em virtude de aplicações das regras de eliminação da disjunção e do quantificador existencial (cf. PRAWITZ, 1965, p. 49; e PRAWITZ, 1971, II.3, p.248, 3.1.2). Há ainda as reduções chamadas *simplificações imediatas,* que visam a remoção de eliminações da disjunção em que não há descarte de hipóteses; também há simplificações imediatas que concernem ao quantificador existencial, e também a aplicações ditas "redundantes" da regra do absurdo clássico.

Graças à analogia descoberta entre a dedução natural e o lambda-cálculo tipado, conhecida como correspondência de Curry-Howard – especialmente efetiva no fragmento conjuntivo-implicativo – é possível verificar que a relação de equivalência obtida a partir do sistema de reduções aqui descrito, em termos do qual a tese da normalização é enunciada, corresponde à βη-equivalência, e à β-equivalência caso as reduções imediatas não sejam incluídas no sistema. Dessa forma, a tese da normalização pode ser formalmente

[13] Aqui seria também plausível questionar por que a segunda redução, em oposição à primeira, insere, ao invés de remover, uma fórmula dispensável para a obtenção da fórmula final de uma derivação a partir das suas iniciais. A resposta, uma vez mais, não se deve a questão da identidade de provas, mas, antes, à *confluência* do processo de redução. Se invertêssemos a direção das expansões imediatas – isto é, se as usássemos como *contrações*, e não expansões –, então, por exemplo, a unicidade da forma normal de derivações em geral seria perdida na presença das reduções por meio das quais se remove fórmulas máximas. Ainda que isso não afete o caso proposicional, ao qual as observações aqui serão restritas, a tese da normalização é enunciada por Prawitz para a lógica de primeira ordem. Novamente, cf. C.-Jay, N.Ghani, *The virtues of eta-expansion.* J. Functional Programming 5 (2): 135-154, April 1995, Cambridge University Press).

considerada como a cláusula de identidade de uma definição de "prova" como um tipo, cuja ideia central é a de que β- e η-conversões – ou, correspondentemente, a de que as conversões respectivamente associadas às reduções que eliminam fórmulas máximas e às expansões imediatas – são preservadoras de identidade.[14]

Deve-se ainda observar que a tese apresenta duas alegações distintas: uma a respeito da *correção*, e outra a respeito da *completude* da relação de equivalência definida em termos das reduções com respeito à relação informal vigente entre derivações que representariam a mesma prova. A cláusula da *correção,* ou a cláusula *se,* diz que a equivalência é *suficiente* para que duas derivações representem a mesma prova, ou, em outras palavras, que a relação formal de equivalência entre derivações é *correta* com respeito à relação informal entre derivações que representariam a mesma prova. A cláusula da *completude*, ou a cláusula *somente se* da tese, por sua vez, diz que a equivalência é *necessária* para que duas derivações representem a mesma prova, isto é, que não há derivações não equivalentes entre si que representem a mesma prova, ou ainda, em outras palavras, que a relação formal de equivalência entre derivações é *completa* com respeito à relação informal entre derivações que representariam a mesma prova.

Longe de um critério arbitrariamente concebido, essa ideia é inspirada em larga medida por concepções filosóficas e resultados formais simples e interessantes, como se pretende deixar claro na sequência.

3. Discutindo a tese: questões formais e conceituais

a. Ideia informal e resultados formais relevantes (possivelmente favoráveis)

Em conformidade com as hipóteses iniciais a respeito de como se manipula as noções de *prova, derivação* e a relação entre elas em um arcabouço semântico, a tese da normalização parece fazer sentido como uma resposta à

[14] Em nome da brevidade, a terminologia do lambda-cálculo será frequentemente empregada em referências às reduções e conversões, ainda que a discussão aqui ocorra no âmbito da dedução natural. Esse dispositivo é especialmente inócuo aqui, uma vez que o fragmento conjuntivo-implicativo é suficiente para quase todas as considerações a serem feitas.

questão da identidade de provas se compreendida como uma expressão formal de uma ideia relativamente simples e plausível, a saber:

> **(α) quaisquer duas provas cujas diferenças entre si se resolvam em traços *irrelevantes* não são significativamente diferentes;**
> **(β) quaisquer duas provas que difiram com respeito a qualquer outro traço são significativamente diferentes.**

A isso, advogados da tese habitualmente acrescentam que

> **(γ) uma prova específica sempre pode ser expressa sem *traços irrelevantes.***

Em suma, a ideia é que quaisquer duas provas são significativamente diferentes se e somente se elas são diferentes *mesmo quando dadas sem quaisquer traços irrelevantes.* As razões e a forma pelas quais as três cláusulas propostas parecem motivar a tese da normalização são discutidas em maior nível de detalhe em de Castro Alves 2018; será oferecida aqui apenas uma indicação de algo de essencial a esse respeito.

A cláusula (α), em primeiro lugar, seria a concepção informal motivadora da parte da *correção* da tese da normalização. Essa alegação se funda fortemente, entre outras razões, numa certa interpretação semântica do chamado *princípio de inversão.* Assim denominado originalmente por Lorenzen, este é formulado por Prawitz em sua tese de 1965 nos seguintes termos:

> Seja α uma aplicação de uma regra de eliminação que tem *B* como consequência. Então, derivações que satisfazem a condição suficiente […] para derivar a premissa maior de α, quando combinadas com derivações das premissas menores de α (caso as haja), já "contêm" uma derivação de *B*; assim, a derivação de *B* pode ser obtida diretamente das derivações mencionadas sem a adição de α."[15] (PRAWITZ, 1965, p.33)

[15] No original: "Let α be an application of an elimination rule that has B as consequence. Then, deductions that satisfy the sufficient condition (…) for deriving the major premiss of α, when combined with deductions of the minor premisses of α (if any), already "contain" a deduction of B; the deduction of B is thus obtainable directly from the given deductions without the addition of α." Tradução livre do autor.

O princípio se refere primeiramente ao seguinte padrão dedutivo:

$$\dfrac{\dfrac{\Pi_1, \ldots, \Pi_n}{*^n(A_1, \ldots, A_{n_1})}\,{*^n}\text{-introdução} \qquad \begin{matrix}\Gamma_1 & & \Gamma_m \\ \vdots & \ldots & \vdots \\ C_1 & & C_m\end{matrix}}{B}\,\alpha$$

Grosso modo, sua alegação é a de que *B,* a conclusão, poderia sempre ser obtida evitando-se a introdução de $*^n(A_1, \ldots, A_n)$ da seguinte maneira:

$$\begin{matrix}\Gamma_i \\ \vdots \\ C_i \\ \vdots \\ \Pi_j \\ \vdots \\ B\end{matrix}$$

Embora isso evidentemente não seja o caso em geral, a restrição às constantes e regras pretendidas pelo autor de fato confirma o princípio. Para cada tal constante, o primeiro padrão dedutivo pode ser reescrito na forma do segundo por meio dos passos de redução descritos por Prawitz na sequência do mesmo texto (p.36 – 38) – todos eles incluídos nos mesmos mencionados aqui na seção 2.1.

Contudo, é importante observar que, logo antes de apresentar os passos de redução, Prawitz refraseia o recém-formulado princípio de inversão em termos mais abertamente semânticos (grifos por minha conta):

> O princípio de inversão diz, com efeito, que nada é 'ganho' inferindo-se uma fórmula por introdução para usá-la como premissa maior em uma eliminação.[16](PRAWITZ, 1965, p. 33 – 34)
> Notamos que a consequência de uma regra-I [regra de introdução] que é também uma premissa maior de uma regra-E [regra de eliminação] constituem uma complicação em uma dedução. Como tal complicação pode ser removida[...], podemos nos perguntar se é possível transformar toda dedução em uma dedução 'normal' correspondente que proceda, por assim dizer, diretamente, sem quaisquer desvios, das hipóteses para a fórmula final.[17](PRAWITZ, 1965, p.34)

[16] No original: "The inversion principle says in effect that nothing is 'gained' by inferring a formula through introduction for use as a major premiss in an elimination." Tradução livre do autor.

Perceba-se, portanto, que os padrões dedutivos removíveis de derivações no decorrer do procedimento de normalização são, desde esse tenro ponto de sua história, caracterizados como, em algum sentido importante, supérfluos. Discutivelmente, a ideia é que esse caráter "supérfluo" se estenderia a todos os padrões dedutivos removidos ou inseridos em derivações em sua normalização, e que estes, em conjunto, instanciariam a noção de "traço irrelevante" em termos da qual a cláusula (α) se enuncia. A parte da completude da tese da normalização equivaleria então à alegação de que *apenas esses* padrões dedutivos instanciariam uma tal noção de "traço irrelevante" – isto é, que quaisquer outros traços quanto aos quais derivações difiram entre si seriam suficientes para estas serem entendidas como significativamente diferentes.

Ademais, o último questionamento de Prawitz no trecho recém-transcrito se deixaria responder afirmativamente pelo chamado *teorema da normalização,* demonstrado pelo próprio na própria tese de 1965[18]. Por razões relativamente simples, esse resultado pode ser convincentemente descrito como o responsável pela adição da cláusula (γ) acima às duas anteriores, e, simultaneamente, como um atestado formal de que, caso as cláusulas (α) e (β) encontrem na tese da normalização uma expressão formal adequada, então a cláusula (γ) se segue delas.

De fato, o teorema da normalização e mais dois resultados formais – a saber, o *teorema da forma normal* e a *unicidade da forma normal* – parecem cruciais, tanto do ponto de vista histórico quanto do conceitual, para a proposta da tese da normalização como um todo.

Em primeiro lugar, o teorema da *forma normal* estabelece que, para toda fórmula *A* e todo conjunto Γ de fórmulas, se *A* pode ser derivada a partir de Γ, então *A* pode ser derivada a partir de Γ de forma *normal* (isto é, sem a

[17] No orignal: "We note that a consequence of an I-rule which is also major premiss of an E-rule constitutes a complication in a deduction. As such a complication can be removed (…), we may ask whether it is possible to transform every deduction to a corresponding 'normal' one which proceeds, so to say, directly, without any detours, from the assumptions to the end-formula." Tradução livre do autor.

[18] Em sua chamada versão fraca, e não contemplando todas as reduções mencionadas na seção 2.1 deste. O teorema da normalização forte, envolvendo todas as reduções mencionadas, é descrito por Prawitz em seu já referido texto de 1971.

ocorrência de fórmulas/segmentos máximos na derivação); isto é, toda relação de consequência válida entre dados *A* e Γ é demonstrável sem recurso a derivações não-normais. Esse resultado pode ser entendido como um tipo de teorema de completude que parece essencial à tese da normalização, que poderia ser enunciado da seguinte maneira: todo *resultado* que pode ser provado pode ser provado de forma *normal*; ou: derivações normais são suficientes para provar todos os *resultados* que podem ser provados.

Além disso, o *teorema da normalização* provê um procedimento mecânico por meio do qual derivações em geral podem ser reduzidas a derivações normais. Isso mostra que a diferença (sintática) entre derivações em geral e suas formas normais se resolve, como já discutido, em certos padrões dedutivos (a saber, aqueles envolvidos nas reduções descritas acima) que "não têm papel" na dedução da fórmula final a partir das fórmulas iniciais não-descarregadas – independentemente da maneira particular em que essa dedução é levada a cabo –; e que supostamente poderiam, portanto, nos termos da discussão acima da leitura semântica do princípio de inversão, ser convenientemente removidos ou inseridos sem consequências para a realização bem sucedida da dedução específica em questão; ou em outras palavras, sem alterar significativamente a dedução em questão. Assim, a cláusula de correção da tese da normalização parece equivaler à alegação de que a remoção/inserção desses padrões dedutivos de/em uma derivação é inócua[19], em algum sentido, para o seu *valor semântico*; e, logo, supostamente irrelevante para a determinação de qual prova é representada pela derivação em que essas remoções/inserções sejam eventualmente realizadas.[20]

A *unicidade da forma normal,* por sua vez, empresta ao procedimento de normalização os contornos de uma *avaliação* em que derivações são julgadas univocamente quanto a qual prova representam, derivações normais assumindo o papel de valores de identidade – ou melhor, em maior conformidade aos termos da formulação oficial da tese, representantes canônicos únicos de

[19] Prawitz reconhece explicitamente no texto de 1971, p. 257, que isso pode não se aplicar no caso dos "redexes" das chamadas reduções permutativas associadas às eliminações da disjunção e do quantificador existencial.

[20] Argumentos para defender essa proposta são explorados na seção III.2.a de de Castro Alves 2018.

provas, que seriam os valores de identidade propriamente ditos. A situação pode ser considerada análoga ao que ocorre com, por exemplo, expressões numéricas em geral e numerais canônicos como representantes de números naturais: por si só, o próprio fato de que "3+2", "4+1", "1+2+2", etc. se reduzem em última análise a "5" e apenas a "5" pode ser entendido como sugerindo que todas essas expressões numéricas teriam o mesmo valor, a saber, 5 – isto é, o valor "direto", "desquotacional" de "5". Dessa forma, derivações em geral podem ser entendidas como *métodos* para a obtenção de derivações normais – exatamente como a expressão numérica "1+2+2" pode ser vista como um *método* (viz. um algoritmo) para a obtenção do numeral "5". Outrossim, assim como "1+2+2" e "5" representam o mesmo número, 5, uma derivação qualquer e a derivação normal à qual esta se reduz representariam a mesma prova. Esse tipo particular de compreensão do procedimento de normalização é presumivelmente uma das pricipais motivações por trás de uma das reformulações mais populares da tese da normalização, usada, por exemplo, por Troelstra[21]: Duas provas correspondentes a derivações Π e Σ são a mesma se e somente se Π e Σ se reduzem à mesma forma normal.

b. Derivações como "representantes" de provas

Uma vez caracterizadas as principais ideias formais e conceituais envolvidas na proposta da tese da normalização, passemos à tentativa de descrição de algumas concepções semânticas mais básicas envolvidas em sua formulação – a saber, as concernentes à ideia de derivações como "representantes" de provas.

A formulação "oficial" da tese carrega consigo um ponto de partida semântico fulcral que constitui, para muito além de uma determinação característica de uma resposta para uma pergunta a respeito da identidade de provas, claras indicações dos termos em que essa pergunta deve ser formulada de saída. Trata-se do seguinte: derivações são concebidas como "representantes" de provas – e estas últimas, em contrapartida, são concebidas como seus valores semânticos. Tomando como ponto de partida a formulação

[21] A.S. Troelstra, Non-extensional equality, Fund. Math. 82 (1975), pp. 307-322.

preliminar da questão da identidade de provas, percebe-se que essa concepção particular dos papéis e da relação semântica entre provas e derivações enseja um deslocamento do que está em jogo nela. Recordando: Dado que Π é uma prova e que Σ é também uma prova, quando é que dizemos que Π e Σ são a mesma prova? – No cenário ora sob consideração, essa pergunta dá lugar àquela para a qual a tese da normalização oferece uma resposta, a saber: "Quando duas derivações representam a mesma prova?". Esta última caracteriza, com efeito, uma busca por como mapear certas entidades linguísticas/sintáticas, as derivações, em valores semânticos, as provas. Note-se que a questão da identidade de provas é assim transfigurada – caso tomemos a questão respondida pela tese da normalização como, de fato, uma interpretação da questão sobre a identidade de provas nos termos da formulação preliminar – de uma busca direta por um critério de identidade entre provas para uma questão que concerne, antes de qualquer outra coisa, à *sinonímia* ou, mais especificamente, à *co-referencialidade* de derivações. Destarte, a identidade dos valores semânticos designados por estas, que nesse cenário, seguindo Prawitz, é o que chamamos propriamente de *provas* (o que quer que se considere que estas sejam), pode permanecer, em princípio, um assunto simplesmente não abordado.

Uma vez estabelecido que a relação entre provas e derivações é a de que estas últimas representam – e assim, em particular, *designam* – aquelas primeiras, algumas conclusões ulteriores podem ser aduzidas levando-se em consideração algumas consequências formais simples da tese da normalização tomada em sua formulação oficial. A primeira a ser notada é a seguinte: dado que há derivações equivalentes distintas, segue-se que uma mesma prova pode ser designada por derivações distintas. Mas o que dizer da conversa desse enunciado? – i.e. poderia uma dada derivação ser tomada por representante de provas distintas? É importante notar que há uma assimetria entre essas expressões das duas direções da relação de representação entre derivações e provas, que se deve à presença de uma ambiguidade na expressão "provas distintas" na formulação da segunda direção, que não ocorre no caso da expressão "derivações distintas". Enquanto "derivações distintas" se refere

neste contexto inequivocamente a quaisquer dois objetos sintáticos distintos entendidos como derivações, "provas distintas" pode querer dizer ou bem, por um lado, provas entre as quais a relação de identidade de provas – independentemente de como se decida entendê-la – não ocorra; ou bem, por outro, provas que sejam simplesmente objetos distintos (numericamente, digamos), a despeito de se esses objetos são ou não identificados pela relação de identidade de provas. Se se toma a expressão no primeiro desses sentidos, parece que a pergunta – a saber: "poderia uma dada derivação ser tomada por representante de provas distintas?" – deveria então ser respondida negativamente no contexto da tese da normalização. Não fosse assim, derivações poderiam ser *ambíguas* quanto à sua referência, i.e. uma única e mesma derivação poderia ter valores semânticos distintos e não equivalentes entre si. E exatamente aqui se evidencia um outro ponto de partida semântico implícito da tese sob discussão: a *univocidade* das derivações com respeito às respectivas provas que representam. No segundo sentido, no entanto, a resposta à pergunta poderia, em princípio, ser positiva. A fim de ilustrar essa colocação, suponha-se que derivações são um *doppelgänger* sintático da prova que designam; exatos "retratos" delas, de tal modo que haveria uma prova distinta correspondente a cada derivação. Dessa maneira, para que uma derivação pudesse designar provas distintas mesmo sob a hipótese da tese da normalização, bastaria que, por exemplo, se admitisse que a relação de identidade viz. equivalência entre os *designata*, as provas, é o exato reflexo daquela entre as derivações que as têm como referência; de tal modo que, digamos, duas derivações distintas Π e Σ que designassem, respectivamente, as consequentemente equivalentes e distintas provas $\mathbf{\Pi}$ e $\mathbf{\Sigma}$ também designariam, respectivamente, $\mathbf{\Sigma}$ e $\mathbf{\Pi}$. Dessa forma, uma derivação poderia se referir a muitas provas distintas – ainda que equivalentes; de fato, a tantas quantas houvesse derivações distintas equivalentes a ela.

Esse segundo cenário parece deformar um pouco a pergunta "quando duas derivações representam *a mesma* prova?" na direção de "quando é que duas derivações representam provas *equivalentes*?"; pois ainda que ela sirva como uma resposta à primeira dessas, ela parece muito mais afinada conforme

a segunda. De qualquer forma, ele goza da virtude de fazer mais justiça à ideia de que a tese da normalização responde, e seus próprios termos, a uma questão que concerne efetivamente à identidade ou à equivalência entre *provas*, ao invés de meramente a alguma relação de equivalência semântica entre derivações; pois ele se constrói em termos de uma certa concepção de prova como uma cópia de uma derivação no nível dos objetos que estas designam. Assim, uma vez que a relação de equivalência entre derivações é estabelecida de um certo modo, também o é aquela da identidade entre provas. Esse pano de fundo semântico para o estudo das propriedades e relações concernentes a provas – e, em particular, da questão da identidade de provas – claramente instancia a estratégia geral descrita por Kreisel em seu Survey of Proof Theory II, de 1971, p.111:

> A natureza geral de nosso problema é bem clara. Considere regras formais cuja função é formalizar certas provas; em outras palavras, temos objetos sintáticos, derivações *d*, que representam ou descrevem atos mentais ***d***, as provas (que trazem convicção); [...] a relação entre *d* e ***d*** […] é um caso particular da relação geral entre palavras e os pensamentos que elas expressam. Como estamos lidando com uma classe "pequena" de palavras, podemos esperar por resultados mais precisos que os conhecidos para a relação mais geral (e mais familiar!). […]
> Dada uma propriedade P ou uma relação R entre provas, nossa tarefa é encontrar relações P_F, resp. R_F, tais que, para toda *d* de nosso sistema formal, P_F (*d*) sse P(***d***) e R_F (*d, d'*) sse R(***d, d'***). Para fins de exposição, inverteremos esse procedimento, e descreveremos primeiro algumas relações formais P_F, R_F que serão então usadas para enunciar os fatos sobre os objetos de interesse principal, a saber, propriedades e relações de provas.[22] [23](KREISEL, 1971, p.111)

[22] No original: "The general nature of our problem is quite clear. Consider formal rules which are intended to formalize certain proofs; in other words, we have syntactic objects, derivations *d* which represent or describe mental acts ***d***, the proofs (which carry conviction);[...] the relation between *d* and ***d*** [...] is a particular case of the general relation between words and the thoughts they express. Since we are dealing with a "small" class of words, we can hope for more precise results than are known for the more general (and more familiar!) relation. (…)
Given a property P or a relation R between proofs our task is to find relations P_F, resp. R_F such that for all d of our formal system P_F (*d*) iff P(***d***) and R_F (*d, d'*) iff R(***d, d'***).
For exposition, we shall reverse this procedure, and first describe some formal relations P_F, R_F which will then be used to state the facts about the objects of principal interest, namely properties and relations of proofs." Tradução livre do autor.

[23] Kreisel parece pressupor, para este tipo de observação, que os enunciados em questão sobre propriedades de provas e derivações são contextos "extensionais"; isto é, em que podemos substituir expressões co-referenciais neles *salva veritate*. Dada a carga epistêmica habitualmente associada ao que tratamos por *provas*, esse não parece um ponto de partida trivial.

Um problema óbvio com essa abordagem dos *doppelgänger* é que ela torna o vocabulário referencialista de "representação" envolvido na formulação da tese da normalização dispensável. A hipóstase de provas como objetos distintos de e designados por derivações se torna, por assim dizer, barba para a navalha de Ockham; afinal de contas, é claramente possível simplesmente tomar a tese da normalização como a caracterização de uma relação de equivalência entre derivações (que poderíamos perfeitamente convencionar chamar de provas – como de fato, aliás, frequentemente se faz), e considerar estas, ao invés de "representantes" sintáticos de objetos de uma natureza mais inefável, como *elas próprias* as provas cuja identidade viz. equivalência se pretende investigar. Assim, seria possível suprimir completamente o modo de expressão referencialista e formular a tese nos seguintes termos: Duas derivações/provas são *sinônimas* se e somente se elas são equivalentes (no sentido de Prawitz). Essa forma de colocar a tese habilitaria a abordagem de provas como expressões linguísticas, e, assim, da questão da identidade de provas como pertinente à *semântica* das expressões linguísticas em questão – na esteira da caracterização do problema a que foi conferida primazia na seção 1.2 acima.

Não obstante, é um fato incontornável que um modo de expressão referencialista *foi usado* na formulação oficial da tese. Assim, para que se possa interpretá-lo de forma mais respeitosa, é mais fácil responder negativamente à pergunta a respeito de se uma derivação pode ou não se referir a provas distintas também no segundo sentido mencionado acima[24], e

[24] Ainda que não seja impossível fazer o contrário. Poder-se-ia, por exemplo, considerar toda derivação β-normal distinta Π_i de *A* a partir de Γ como um representante canônico de uma prova $\mathbf{\Pi_i}$ correspondentemente distinta, e dizer que toda derivação que se reduz a Π_i também designaria a prova $\mathbf{\Pi_i}$. Além disso, sejam as $\mathbf{\Pi_i}$, a despeito de serem distintas entre si, todas *equivalentes* entre si; digamos, por exemplo, pelo fato de todos os seus representantes canônicos serem em última análise redutíveis a uma única forma βη-normal. Então, toda derivação que se reduzisse a Π_i designaria a prova $\mathbf{\Pi_i}$, mas não apenas ela; elas também designariam as provas equivalentes a $\mathbf{\Pi_i}$. Poder-se-ia então interpretar o sentido desse cenário da seguinte maneira: derivações β-normais seriam os *doppelgängers,* os representantes canônicos de provas concebidas como os conteúdos *intensionais* de derivações, de tal modo que provas β-conversíveis entre si compartilhariam esse conteúdo intensional. Estes últimos, as provas, seriam *eles próprios* equivalentes/idênticos quando seus representantes canônicos se reduzissem a uma mesma forma βη-normal. Nesse cenário, derivações podem designar mais de uma prova; as provas não podem ser simplesmente consideradas barba para a navalha de Ockham sem mais; e a abordagem ofereceria uma resposta não apenas à questão da equivalência semântica entre derivações, mas também, em seus próprios termos e literalmente,

entender a tese da normalização como uma solução para a questão da identidade de provas de uma forma bastante simples. A saber: primeiramente, como sugerindo de saída que cada derivação, além de *unívoca*, i.e. além de *referir-se apenas a objetos equivalentes*, também se refere a *apenas um* objeto (prova); em segundo lugar, como sugerindo que a relação de equivalência entre derivações é, tal como observado, necessária e suficientemente condicionada pela sua *co-referencialidade* – isto é, pelo fato de que designam *uma única e mesma prova*; e terceiro, como sugerindo que a relação de identidade/equivalência concernente às *provas* elas próprias, os objetos designados, deve ser tomada por vigente somente entre um objeto e ele mesmo, i.e. é trivial e, por conseguinte, desinteressante. Ainda que essa abordagem torne difícil compreender em que sentido a tese da normalização responde a algum problema digno do rótulo "identidade de provas" – ao invés de, digamos, co-referencialidade de derivações –, ele certamente acomoda muito melhor o modo referencialista de expressão adotado na formulação da tese.

Neste ponto da reflexão, em que já se evidenciou que não é em absoluto necessário hipostasiar provas como *designata* de derivações sintaticamente concebidas para se formular a tese da normalização de modo plausível; e também que, com efeito, a hipóstase de tais objetos e relações semânticas referenciais torna a tese bastante suscetível a acusações sérias, como a de postular existências sem propósito ou simplesmente não abordar um problema digno do título "identidade de provas"; caberia questionar quais seriam então as razões para escolher formular a tese da normalização em seu registro referencialista tradicional. A respeito desse assunto, me parece particularmente interessante sublinhar que a tese da normalização, tal como enunciada por Prawitz em 1971, é imbuída de contornos *descritivos*, e não *estipulativos* – de fato, é frequente na literatura referir-se à tese da normalização como a "conjectura" da normalização; fê-lo o próprio Prawitz ao enunciá-la, além de, por exemplo, Kosta Došen 2004 e Filip Widebäck 2001. Em outras palavras, a

àquela a respeito da identidade de provas. Uma tal teoria precisaria, contudo, de fundamentação filosófica sólida a fim de não se reduzir a um mero dispositivo *ad hoc* em prol da defesa, simultaneamente, da relevância da hipóstase de provas demandada pela formulação representacionalista da tese da normalização, e de que o assunto abordado pela tese da normalização é de fato a identidade de provas, e não apenas a equivalência entre derivações.

tese é enunciada como uma proposta possivelmente verdadeira, possivelmente falsa, de descrição da relação de identidade de provas, e não como uma definição que, mais do que meramente descrever, *estipula* a natureza de tal relação. Esse interessante e importante aspecto da tese da normalização é amplamente discutido no capítulo V de de Castro Alves 2018, donde se pode extrair material para uma compreensão mais sólida da diferença que ele faz para o ponto ora sob consideração. Aqui, restringir-se-á as observações a esse respeito ao seguinte: a hipóstase de provas como objetos designados por derivações permite uma solução de espírito fregeano para a questão de se duas derivações representam ou não a mesma prova. Assim como a estrela da manhã e a estrela da tarde são idênticas em virtude de serem o mesmo corpo celeste, ou 117+136 e 253 são idênticos em virtude de serem o mesmo número natural – o que implica que "a estrela da manhã" ["117+136"] e "a estrela da tarde" ["253"], a despeito de suas eventuais diferenças em *Sinn* (sentido), são des-crições do mesmo objeto (i.e. têm a mesma *Bedeutung* (significado, referência)) e podem, assim, ser argumentos de um enunciado de identidade verdadeiro e informativo –, também seria o caso de, digamos,

$$\dfrac{\dfrac{\Pi_1 \quad \Pi_2}{A \quad B}}{\dfrac{A \wedge B}{A}}\ \Pi_3 \quad \text{e} \quad \begin{matrix}\Pi_1\\ A\\ \Pi_3\end{matrix}$$

(e aqui as derivações estão sendo *usadas*, e não mencionadas) serem idênticas em virtude de serem a mesma prova – o que implica que

$$\text{“}\ \dfrac{\dfrac{\Pi_1 \quad \Pi_2}{A \quad B}}{\dfrac{A \wedge B}{A}}\ \Pi_3\ \text{“} \quad \text{e} \quad \text{“}\ \begin{matrix}\Pi_1\\ A\\ \Pi_3\end{matrix}\ \text{“}$$

(e agora as derivações estão sendo *mencionadas*), apesar de seus "sentidos" eventualmente diferentes, são "representantes" do mesmo objeto, servindo assim como argumentos para um enunciado de identidade verdadeiro. O modo

de expressão referencialista pode então ser compreendido como um ingrediente essencial à possibilidade de se recorrer à idéia de que *provas*, tomadas como efetivos *objetos*, e sua natureza, i.e. as propriedades que apresentam e as relações de que participam, são o derradeiro fundamento para a (in)adequação de um critério para a verdade ou falsidade de enunciados a respeito da equivalência semântica de derivações – o que seria, de outro modo, inteiramente dependente de estipulações ou convenções a respeito do sentido (*Sinn*, ou – para nos desvincular das terminologia e doutrina fregeanas – significado em acepção *intensional*) dos últimos, relegando a tese da normalização ao estatuto de uma abordagem puramente definicional da sinonímia/identidade de derivações/provas. Como já observado, contudo, o rótulo "identidade de provas" para uma investigação concebida em tais termos segue sendo relativamente enganoso: assim como a investigação de Frege em *Über Sinn und Bedeutung* não toma por objeto a identidade de corpos celestiais, números naturais ou outros objetos, mas é antes "*über Sinn und Bedeutung"* – i.e. concerne a fenômenos relacionados à semântica de *expressões linguísticas* que *se referem a* corpos celestes, números naturais e outros objetos –, a ora sob escrutínio tampouco parece concernir à identidade dos *designata* chamados de provas, mas, antes, à co-referencialidade de derivações concebidas como objetos sintáticos. Ademais, há ainda um "elefante na sala" caso se queira adotar essa abordagem referencialista à questão da identidade de provas: se, por um lado, sabemos muito bem o que são, por exemplo, corpos celestes e como identificá-los e distinguí-los adequadamente, não parece que temos o mesmo grau de clareza a respeito dos elusivos supostos objetos que nesse contexto chamamos de provas.

4. Conclusão

Além de uma caracterização da questão da identidade de provas e de sua legitimidade e relevância como um problema filosófico, o presente esforço debruçou-se também sobre da tese da normalização, a principal proposta de abordagem de índole formal desse problema filosófico. Procurou-se não apenas descrever os aspectos explícitos desta última, como também evidenciar

certos pressupostos semânticos basilares e críticos de sua formulação tradicional, analisando também sua influência no grau de plausibilidade da tese.

Observou-se que os termos referencialistas em que a tese da normalização é habitualmente formulada põem em xeque, de acordo com sua própria economia terminológica, a possibilidade de a tese oferecer, mesmo em princípio, uma resposta que concirna propriamente à identidade de provas. Ademais, a caracterização de derivações como meros "objetos sintáticos" – de aroma metamatemático e tributária de uma concepção de linguagem a partir da cisão forçada e, após Tarski, hegemônica e naturalizada em lógica, entre *sintaxe* e *semântica* – dificulta que se as compreenda como compatíveis com análises semânticas de qualquer natureza independentemente de atribuir-se a elas dimensões semânticas que, por assim dizer, as transcendam.

Buscou-se, contudo, evidenciar também que o critério de equivalência semântica entre derivações em jogo na tese da normalização não parece depender dos termos representacionalistas de sua formulação habitual, e parece, assim, defensável em princípio como uma proposta de resposta à questão da identidade de provas como concernente à sinonímia de provas (derivações) tomadas como expressões linguísticas.

Maiores reflexões acerca do grau de plausibilidade da tese da normalização tomada em seus diferentes registros – referencialista ou não referencialista, descritivo ou estipulativo – são levadas a cabo em de Castro Alves 2018. Para os propósitos deste esforço, é cabível apenas repetir o que Kreisel já em 1971 dissera a respeito dessa abordagem do problema aqui apresentado, e que permanece verdadeiro a despeito de seus (de)méritos ulteriores: "no presente momento, o critério [da tese de normalização] tem uma utilidade pedagógica: ele corrige a pressuposição comum de que "nada" preciso (e razoável) pode ser feito a respeito de questões sobre sinonímia de provas."[25] (KREISEL, 1971, p.117)

[25] No original: "...at the present time, the criterion [of the normalisation thesis] has a pedagogic use: it corrects the common assumption that "nothing" precise (and reasonable) can be done on questions about synonymity of proofs." Tradução livre do autor.

5. Referências

ARISTÓTELES. *Aristotle's Prior Analytics book I*. Translated with an introduction and commentary by Gisela Striker (Clarendon Aristotle Series). New York: Oxford University Press, 2009.

DE CASTRO ALVES, Tiago. *Synonymy and Identity of Ṕroofs: a Philosophical Essay*. Tese (Doutorado em Filosofia) – Eberhard-Karls-Universität Tübingen, Tübingen, 2018.

DOŠEN, Kosta. "Identity of proofs based on normalization and generality". *The Bulletin of Symbolic Logic*, vol. 9, 2003, p. 477-50.

FREGE, Gottlob. "Über Sinn und Bedeutung". *Zeitschrift für Philosophie und philosophische Kritik*, NF 100, 1892, p. 25-50

JAY, C. Barry, e GHANI, Neil. "The virtues of eta-expansion." *J. Functional Programming*, 5 (2), 1995, p. 135-154

KANT, Immanuel. *Escritos Pré-Críticos.* Trad. Jair Barboza et al. São Paulo: Edunesp, 2005.

KANT, Immanuel. *Lógica.* Trad. por Guido Antônio de Almeida. (Série Estudos Alemães, 93). 2ª ed. Rio de Janeiro: Tempo Brasileiro, 1992.

KREISEL, Georg. "A survey of proof theory II". In: FENSTAD, J.E. (ed.), *Proceedings of the Second Scandinavian Logic Symposium.* Amsterdam: North-Holland, 1971, p. 109-170.

PRAWITZ, Dag. *Natural Deduction. A Proof-Theoretical Study*. Stockholm: Almqvist & Wiksell, 1965.

PRAWITZ, Dag. "Ideas and results in proof theory". In: FENSTAD, J.E. (ed.), *Proceedings of the Second Scandinavian Logic Symposium.* Amsterdam: North-Holland, 1971, p. 235-307.

SCHROEDER-HEISTER, Peter. "Proof-Theoretic Semantics", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/proof-theoretic-semantics/>

SUNDHOLM, Göran. "Identity: Absolute. Criterial. Propositional". In: CHILDERS, T. (ed.), *The Logica Yearbook 1998*. Praga: The Institute of Philosophy, Academy of Sciences of the Czech Republic, 1999, pp. 20-26.

TROELSTRA, A. Sjerp. "Non-extensional equality". *Fund. Math.* 82, 1975, p. 307-322.

WIDEBÄCK, Filip. *Identity of Proofs.* Tese (Doutorado em Filosofia) – Stockholms Universitet, Almqvist & Wiksell, Stockholm, 2001.

INTEGRAÇÃO DO *INCREASED COHERENCE MODEL* DE GALLIERS À MODELAGEM EVANS-ARAÚJO DO JOGO DE DAR E PEDIR RAZÕES DE ROBERT BRANDOM

Ralph Leal Heck[1]

RESUMO

O objetivo deste artigo é reavaliar uma formalização do jogo de dar e pedir razões de Brandom (1998) proposta por mim em (HECK, 2018), que a divide em duas instâncias: uma individual de performances potencialmente públicas e outra efetivamente histórico-pública; e propor correções à primeira das duas instâncias. Nomeadamente, substituindo a formalização em banco de dados e operações estáticas, elaborada por Araújo (2018), pela expressão formal de uma atividade cognitiva e comunicativa dinâmica dada por Galliers (1992) chamada de ICM (*increased coherence model*) a partir do paradigma AGM (1988). Para realizar isto, inicio apresentando o pensamento de Brandom e sua concepção de jogo de dar e pedir razões, com suas dinâmicas subjacentes, em seguida apresento uma formalização que combina a instância pública de Evans (2018) com a instância pessoal de Araújo. A partir daí, elenco algumas críticas, com especial foco no caráter injustificado e estático das operações sobre os estados doxásticos do modelo de Araújo. Por fim, apresento a modelagem de Galliers como solução para estas críticas, concluindo que esta incursão de revisão de modelagens adiciona elementos pragmáticos às operações sobre crenças, embora, apenas o início das revisões necessárias para fornecer uma expressão lógica satisfatória da atividade dinâmica e inconsistente que é o jogo de dar e pedir razões.

Palavras-chave: Revisão de Crença. Paradigma AGM. Robert Brandom. Increased Coherence Model. Jogo de Dar e Pedir Razões. Raciocínio Não-Monotônico.

ABSTRACT

The objective of this article is to reevaluate a formalization of Brandom's (1998) game of giving and asking for reasons proposed by me in (HECK, 2018), which separates it into two instances: an individual that can be turned into public performances and another which is effectively historical-public; proposing corrections to the first of these instances. Namely, replacing the formalization in database and static operations, elaborated by Araújo (2018),

[1] Doutor em Filosofia, professor de Filosofia pela Universidade Federal do Ceará e pela Faculdade Católica de Fortaleza.
E-mail: imagomundi@hotmail.com. ORCID: 0000-0002-9827-1743.

by the formal expression of a dynamic cognitive and communicative activity given by Galliers (1992) called ICM (increased coherence model) from the AGM paradigm (1988). To accomplish this, I begin by presenting Brandom's thinking and his conception of giving and asking for reasons, with their underlying dynamics. Then I present a formalization which combines the public instance of Evans (2018) with the personal instance of Araújo. From there, some criticisms being cast, with special focus on the unjustified and static character of operations on the doxastic states of Araújo's model. Finally, I present Galliers' modeling as a solution to these criticisms, concluding that this incursion of modeling revision adds pragmatic elements to operations on beliefs, although, being only the beginning of necessary revisions to provide a satisfactory logical expression of the dynamic and inconsistent activity that is the game of giving and asking for reasons.

Keywords: Belief Revision. AGM Paradigm. Robert Brandom. Increased Coherence Model. Game of Giving and Asking for Reasons. Nonmonotonic Reasoning.

1. Sobre o jogo de dar e pedir razões de Brandom

A obra *Making it Explicit* de Robert Brandom (1998), tem como cerne de sua filosofia da linguagem pragmático-inferencialista o conceito de *jogo de dar e pedir razões*. Isto significa dizer que o conteúdo de nossos conceitos é construído na atividade coletiva de inferir conceitos a partir de outros e reconhecer os conceitos dos quais podemos inferi-lo. Para Brandom, a semântica e a própria racionalidade repousam na atividade de inferir e expressar estas inferências. O significado (o conteúdo semântico) de uma sentença é dado pelas inferências feitas para se chegar a ela e pelas que fazemos a partir dela. Onde cada sentença, expressa em asserções, tem seu significado determinado enquanto partícipe de inferências. O modo de determinar esta participação é pelo *uso* das sentenças nas inferências. A própria noção de atribuição sentencial (ou proposicional - intencional) é universalmente aplicável, na medida em que esta atribuição feita pelos usuários da linguagem, no interior de inferências, fornece razões pelas quais algo ou alguém fez, faz ou fará alguma coisa. As asserções se fundam nos componentes subsentenciais que são definidos do mesmo modo: o significado de um conceito é determinado por seu papel inferencial, na medida em que, damos justificações para sua aplicação.

A totalidade desta prática de inferências e asserções é o que nos distingue de simples papagaios (BRANDOM, 1998. p.89), é a atividade proficiente nesta aplicação de conceitos que chamamos de *jogo de dar e pedir razões*. E saber jogar este jogo constitui a própria racionalidade humana, na medida em que, ela é sinônimo de expressabilidade, isto é, a racionalidade humana é a capacidade de tornar explícito aquilo que está implícito na totalidade de nossas ações. Nestes termos, o sentido pragmático de sua filosofia é a afirmação de que a inferência é um *fazer*, porquanto é a habilidade racional de operar asserções em um percurso inferencial, regulado por regras, realizando e tornando todos estes itens explícitos no interior da prática social de comunicação. A partir desta prática entre os indivíduos (as asserções e inferências compartilhadas entre eles), a própria racionalidade é construída no curso da expressão do que se faz e na interação mútua com o que é expresso, portanto, do que é feito por outros indivíduos (ou coisas).

Mas, de que forma podemos compreender a estrutura deste jogo de dar e pedir razões? Da perspectiva semântica, trata-se de uma compreensão holística do significado dos conceitos, eles estão vinculados por inferências e incompatibilidades entre elas, por exemplo, se digo que estou no hemisfério sul, então não estou autorizado a afirmar que estou no hemisfério norte, mas estou autorizado a dizer que estou no Brasil, se houver argumentos adicionais que endossem isto. Se dissesse inicialmente que estou no Brasil, estaria comprometido com a afirmação de que estou no hemisfério sul.

De uma perspectiva pragmática, o jogo é uma atividade coletiva em que cada participante desempenha um duplo papel: (i) de interlocutor, aquele que possui crenças expressas na forma de asserções, faz perguntas, responde perguntas e (ii) de contador de escores (*scorekeeper*) que é ser uma espécie de fiscal das ações dos interlocutores no decurso de um diálogo, no intuito de acompanhar a correção das ações em função das regras constitutivas do contexto de comunicação. Isto é, tomamos conta de si e dos outros conferindo se as respostas estão de acordo com as regras subjacentes às inferências em questão.

A correção do jogo é determinada materialmente (em oposição à noção de construção formal), ou seja, as regras são constituídas materialmente (onde a lógica entra com o papel de explicitar, não de normatizar sobre quais inferências são válidas – o que é chamado por Brandom de *expressivismo lógico*). Ser constituído materialmente significa dizer que o vínculo inferencial entre conceitos e entre uma asserção e a próxima é determinado histórico-culturalmente por regras que possuem caráter normativo. As inferências lógicas e suas regras associadas são apenas uma classe dessas inferências (também normativas), o padrão de funcionamento destas classes é determinado a partir de um vocabulário fixado (que identifica o domínio do discurso) e de regras próprias ao domínio em questão, considerando sua natureza contextual e normativa. O que significa dizer que podemos falar de inferências médicas, químicas, teológicas, poéticas, políticas, etc. posto que tenhamos fixamos o vocabulário do domínio e identificados as inferências materialmente válidas para aquele vocabulário (*Ibid.* p.104).

As asserções constituintes da prática de comunicação, na medida em que são performances identificadas como apropriadas e inapropriadas mediante normas, possuem duas propriedades: as circunstâncias apropriadas de aplicação de uma asserção e as consequências apropriadas desta aplicação (*Ibid.*. p.159). À primeira noção, chamamos de autorização e à segunda de compromisso. Desta forma, ao realizarmos uma asserção, devemos levar em consideração quais outras asserções estou habilitado a fazer em seguida e com quais outras asserções estou comprometido e, eventualmente, impedido explicita e implicitamente.

Nestes termos, a intencionalidade é definida e articulada linguisticamente por meio destes *status* deônticos e, se a condição básica de nossas atitudes proposicionais e nossos atos de fala são determinados na atividade da linguagem, então nossa mente também o é (*Ibid.* p.152), assim, acreditar e afirmar (*belief and claim*) estão relacionados essencialmente, de forma que o primeiro é explicado em termos do segundo (*Ibid.* pp.154-155). Por isto, as crenças podem ser modeladas como um tipo de compromisso inferencialmente articulado, chamado por Brandom de *compromisso doxástico* ou assercional

(*Ibid.* p.157) e este é o tipo de compromisso discursivo mais básico dos falantes. A expressão *compromisso doxástico* deve ser entendida como o vínculo do falante com um conteúdo asserível, isto é, o papel inferencial que o conteúdo conceitual de uma sentença declarativa desempenha. Por causa disto, nossos compromissos doxásticos, como o nome diz, nos comprometem com certo conteúdo e, a partir daí, nos encontramos autorizados, comprometidos e desautorizados por meio de regras normativas a asserir outros conteúdos. Caso o indivíduo insista em sustentar compromissos doxásticos desautorizados, apesar das regras que o constrange, ele passa a apresentar *crenças inconsistentes* (*Ibid.* p.160). Uma situação deste tipo evoca sobre o falante uma força normativa de correção deste estado inconsistente (manifesta por si mesmo ou por outros falantes).

As regras que medeiam estas passagens ou impedimentos de asserções, por seu turno, são expressas como condicionais subjuntivas (*subjunctive conditionals*). Elas habilitam as transições inferenciais entre asserções e a expressão (asserção) da própria regra, tornando explicito o que está implícito em nossos raciocínios. Como são constituídas materialmente e trazem consigo um valor deôntico, as regras inferenciais são divididas em regras de autorização, compromisso e incompatibilidade.

As regras de autorização dizem respeito ao que posso afirmar em face de asserções (crenças – estados doxásticos) que já tenham sido proferidas ou estejam disponíveis no curso da comunicação. Por exemplo, se afirmo que ontem, às 9 horas da manhã, o mercado estava aberto, e que outras pessoas participantes de dado diálogo também viram que o mercado estava aberto às 9 horas da manhã em outros dias, logo, se afirmo que são 9 horas da manhã, então estou autorizado a asserir que o mercado está aberto. Uma classe paradigmática deste tipo de inferências, como o exemplo mostra, são as inferências indutivas. Já as inferências de compromisso são aquelas com as quais, a partir de uma asserção, comprometo-me com uma série de outras sentenças. Uma classe de inferências características deste tipo são as deduções. Por exemplo, se me comprometo com a afirmação de que “sempre que chove, as ruas ficam

molhadas" e com a afirmação de que "hoje está chovendo", então estou comprometido com a sentença "as ruas estão molhadas"[2].

Já as relações de incompatibilidade são aquelas que determinam os casos incompatíveis em função de uma asserção (as incompatibilidades são usualmente vinculadas a expressões contrafactuais). Por exemplo, se afirmo que a velocidade de queda de um corpo na Terra é de aproximadamente 9,8 metros por segundo ao quadrado, então, se afirmasse que a velocidade de queda é 5 metros por segundo ao quadrado, não estaria falando da terra. Pois a afirmação do valor 9,8 atribuído à Terra exclui todas as outras possibilidades, salvo entrar em contradição com o compromisso anterior.

Para dar movimento a este jogo de asserções e relações inferenciais no curso dos diálogos (com um ou mais interlocutores), considerarei duas instâncias de construção inferencial feitas em (HECK, 2018. 96ss; 105ss): 1) Os sujeitos tomados em si mesmos, como entidades autônomas, trabalhando em termos de inferências acerca do que já sabem e o que vêm a saber e 2) as regras materiais que governam as inferências produtoras de outras asserções. Tanto o acompanhamento das performances, como o reconhecimento das regras é algo maleável. Por exemplo, é possível tomar para si uma asserção de outro indivíduo (neste caso, comprometendo-se com o que se segue da asserção) e também adquirir, corrigir ou abandonar as regras que orientam certa etapa do discurso. Evidenciando o caráter "interpessoal, intrapessoal, intraconteúdo e interconteúdo dos compromissos", de tal forma que, esta estrutura discursiva combine a autoridade pessoal, na medida em que, podemos confrontar uma asserção com outra e a autoridade do conteúdo, ao justificarmos nossa asserção, tornando explicito seu percurso inferencial realizado pelos indivíduos e as regras usadas para se chegar a ela (BRANDOM, 1998. p.175).

Pelo que vimos, as asserções dos sujeitos devem estar em conformidade com as regras inferenciais ao longo do diálogo. Mas, não é o caso que os sujeitos saibam de todas as regras (eles não são materialmente oniscientes), ou que tenham absoluta consciência da consistência de suas crenças e da condução das asserções por todo o curso da comunicação (nem logicamente onisci-

[2] O exemplo expressa uma dedução materialmente válida.

entes). Brandom estabelece algumas funções diretivas e corretivas para o jogo de dar e pedir razões, a partir da noção de *scorekeeping* (*Ibid.* pp.181-182)[3]. Neste sentido, o jogo se assemelha ao xadrez. A cada lance de um jogador A, o jogador B (ou outros jogadores) têm ações, antes possíveis, agora impossibilitadas, além de serem capazes de antever alguns lances disponíveis para A e, igualmente, os indisponíveis na próxima passagem de turno do jogo. Em acréscimo a isto, quando um enxadrista (iniciante) comete um erro de movimento, podemos pedir que ele repare a jogada com uma nova, agora correta, segundo as regras do xadrez. Do mesmo modo, quando um falante assere algo que não está de acordo com as regras, ou que não seria permitido em função de alguma incompatibilidade, os interlocutores, podem solicitar que o falante dê razões para sua asserção ou corrija sua asserção explicitando a regra que a constringe, ou ambos os casos.

2. Incursão na formalização do jogo de dar e pedir razões

Em (HECK, 2018. pp.99-100) propus um modo de formalizar este jogo de dar e pedir razões combinando dois modelos distintos, tais que operem como dois parâmetros de funcionamento do jogo. O primeiro, compreendido como pessoal, contendo as crenças (compromissos doxásticos) expressáveis em performances, as regras de inferência conhecidas e as competências cognitivas do indivíduo. O segundo, público, que registra o estado atual da comunicação com base nas regras evocadas até o momento. Como se registrássemos a visão atual do tabuleiro de xadrez acompanhado do histórico de jogadas dos enxadristas.

Ao primeiro modelo, atribuí uma representação em termos de banco de dados de cada falante contendo seus compromissos doxásticos e as regras de inferência conhecidas e aplicáveis ao contexto de fala, além de um conjunto de operações sobre os compromissos doxásticos (onde as regras são estáticas). Ao segundo modelo, atribuiu-se a formalização do aspecto público das

[3] Para uma explicação da prática de contar escores como regras e estados de transição de asserções entre agentes (cf. HECK, 2018. pp.96-98). Não entraremos nestes detalhes, uma vez que não parecem problemáticos à luz de Brandom, nem à de Lewis (cf. LEWIS, 2004.) e nem à luz da linguagem de autômatos.

performances dos falantes no contexto de prática comunicativa, chamado de estado de debate (EVANS, 2016. p.395), de tal forma que, no modelo estão incluídas regras inferenciais de autorização e incompatibilidade. No primeiro modelo, a formalização foi executada seguindo o trabalho de Araújo (2016. p.129-140.), o segundo modelo foi construído com base no trabalho de Evans, que por sua vez, modela o jogo de dar e pedir razões (EVANS, 2016. p.395ss) a seu modo, focalizado no aspecto das regras e incompatibilidades.

2.1 A modelagem Evans - Araújo

O estado de debate (D_X), que demarca o contexto de comunicação, é definido por uma dupla <X, S>, onde $x \in X$ são os falantes e S são as sentenças disponíveis em uma linguagem L em uso pelos falantes. Este contexto é definido por uma quadrupla A, C, E, I, tal que A é um conjunto de duplas contendo asserções registradas e publicamente atribuídas a um dos interlocutores (e.g. o registro de cada movimento feito pelos jogadores de xadrez): $A \subseteq X \times S$; C é o conjunto de inferências de compromisso e E é o conjunto de inferências de autorização (*entitlement*), tais que $C \subseteq P(S) \times S$ *e* $E \subseteq P(S) \times S$. É importante notar que Evans confere às inferências de compromisso e autorização o caráter de cláusulas Horn[4].

Uma vez que a quadrupla A, C, E, I modela a situação de debate no qual participam indivíduos, Evans sugere uma modelagem das autorizações específicas destes indivíduos como um subconjunto da situação de debate: $Entitled_x(A, C, E, I) \subseteq A$ (p.396). Uma vez que tal estrutura captura as autorizações de x, parece razoável sugerir que podemos circunscrever os lances totais de x e as regras associadas ao longo do debate incluindo as inferências de compromissos. E, além disso, acrescentar os processos inferenciais na forma de compromissos doxásticos que ele realiza antes e após cada lance.

---

[4] Onde P(S) é a cauda da cláusula e S é a cabeça da cláusula e cada cláusula Horn é representada pela expressão $(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n) \rightarrow \beta$ e $P(S) \times S$ é descrito como $(\psi_1, \psi_2, \ldots, \psi_n) \in P(S), \phi \in S$. Já as relações de incompatibilidade de cláusulas Horn são expressas como conjuntos de conjuntos de sentenças incompatíveis $I \subseteq P(P(S))$= $Inc(S) = \{x \in S \mid \exists\, y_1, \ldots, y_n \in S \text{ tal que } \{x, y_1, \ldots, y_n\} \in I\}$.

Para operar no interior dos compromissos doxásticos de modo a combiná-los em nexos mais complexos e capturar a atividade de aprendizagem, correção de crenças e remoção delas, acrescentei à quadrupla de Evans uma dupla $B \times Op(B)$ desenvolvida por Araújo com o intuito de capturar a informatividade contida em revisões de crenças (cf. HECK, 2018. pp.105-106). Onde B é uma dupla $\mathring{A}, T$ tal que $\mathring{A}$ seja uma estrutura finitária em linguagem de primeira ordem com uma função de interpretação R, expressando o vocabulário e as sentenças atômicas portadas pelo indivíduo x e T um conjunto de sentenças fechado sobre a relação de consequência lógica (clássica), tendo $\mathring{A}$ como premissas básicas e $Op(B)$ como operações de inserção e deleção sobre a base de crenças $\mathring{A}$. Estas operações buscam modificar $\mathring{A}$, seja no vocabulário, seja na lista de sentenças atômicas.

Tomemos a interpretação de T como "o conjunto de compromissos e regras de inferência material adotado pelo agente" no contexto de comunicação. Tal que, $B \times Op(B)$ seja lido como o estado atual do falante neste contexto e T seja um conjunto de expressões na forma de cláusulas Horn, modificável quando o caso de alguma incompatibilidade prescrita. Assim, passa-se a expressar num contexto de Debate D_x as autorizações e compromissos de x como uma quíntupla: $Entitled \wedge Committed_x$ A, T , E, I, B, Op (B/ T /E,

em que A são as asserções já proferidas, T é o conjunto de inferências de compromisso, E o conjunto de inferências de autorização, I o conjunto de conjuntos de incompatibilidade, B os estados doxásticos de x relevantes em D_X e suas operações OP($B/T/E$), além de aplicáveis a B, também são aplicáveis à T e E. As operações sobre T e E, por sua vez, correspondem analogamente às operações de inserção e deleção, mas aplicáveis às cláusulas Horn.

2.2. Críticas à estratégia Evans-Araújo

Embora esta combinação de formalizações forneça um modo de expressar a dinâmica do jogo de dar e pedir razões de Brandom, cada uma das estratégias deixa lacunas a seu modo. No caso de Araújo (2016. 130ss.), o banco de dados descrito pela dupla $B \times Op(B)$ deixa implícita as operações de

composição de sentenças complexas. Também não há mecanismos de identificação de pesos epistêmicos de sentenças e as operações lógicas que produzem sentenças são consideradas primitivas servindo na identificação de informatividade, com isto, seu sistema não captura o aspecto material das inferências que governam nosso raciocínio.

Dito de outro modo, o que justificam as operações? A estrutura descrita por Araújo não lida com aspectos epistêmicos e pragmáticos fundamentais para Brandom. Por isso, por exemplo, adicionei uma instância de regras materiais à dupla $\mathring{A}, T$. Mas, tais regras não são justificadas internamente ao sistema. Do mesmo modo, a qualidade dos *inputs* do sistema não é posta em questão, não há como diferenciar o caráter social, emocional, etc. das asserções que assumimos para nós. Por exemplo, em Brandom, as asserções que endossamos de terceiros e a confiança na responsividade de outras entidades não é classificado monotonicamente (cf. BRANDOM, 2000. p.88), esta não-monotonicidade não deve ser expressa apenas na maleabilidade das regras, mas na própria forma como hierarquizamos e assimilamos as asserções e compromissos doxásticos de terceiros. E, por causa disso, é provável e cotidianamente ocorrente que um indivíduo assimile crenças incompatíveis e depois se coloque na situação de ter de escolher em qual delas confiar.

A crítica que reservo a Evans pertence à mesma natureza das últimas observações acima, ela diz respeito à pressuposição de consistência entre as regras constitutivas do debate. Ora, é possível apontar regras histórica e publicamente constituídas que sejam incompatíveis. No texto de Evans e também no meu não ficou claro como se expressam estas reparações de regras de inferência. Se por um lado, Evans apenas fornece uma estratégia para evitar a parada no fluxo do debate, acrescentando sempre novas sentenças, por outro, me propus a tratar as cláusulas Horn a partir do paradigma AGM. Mas, por mais que o paradigma seja eficiente para manipular as cláusulas, refletindo aspectos epistemológicos fundamentais, sua modelagem não lida com incompatibilidades efetivas, apenas em casos de divergência entre a classe de inferências assimiladas pelo sujeito e a classe de inferências públicas. Estas últimas pressupostas como consistentes. Mas, com lidar com casos de inconsistência

real (persistente) entre as próprias regras constitutivas? Isto deve significar uma parada no debate? Este não parece ser o caso se olharmos para o cotidiano, das conversas de bar e das postagens de internet às audiências legais e modelos da física.

Para tratar os problemas associados à formalização de Araújo sugiro substituir todo seu aparato formal pelo aparato do paradigma AGM, mais especificamente, à modelagem proposta por Galliers (GALLIERS, 1992. p.220), que procura assimilar na própria organização dos estados doxásticos a situação aberta e dinâmica de operações sobre estes estados no contexto de comunicação entre indivíduos. Além do aspecto dinâmico, ela introduz nesta instância uma segunda camada de raciocínio não-monotônico e pragmático. A primeira já encontrada em Araújo corresponde à não-monotonicidade temporal e é retratada pelas operações OP sobre B, mas a segunda, não tocada por ele, corresponde à não-monotonicidade lógica[5] e significa o tratamento sobre os resultados inconsistentes de inferências obtidas como produto de raciocínios corretos.

3. A modelagem de Galliers

A proposta para solucionar isto é justamente conduzir da modelagem de Araújo para a modelagem de Galliers, que trabalha no paradigma AGM. Em certo sentido, trazendo para um nível mais básico a modelagem das regras.

No decurso de nossas experiências, mudamos de crença acerca daquilo que é ou não o caso. Aprender uma coisa nova, estar errado sobre algo e corrigir-se ou simplesmente deixar para lá uma coisa que se pensava conhecer são situações que envolvem fundamentalmente a revisão de nossas crenças. A preocupação com estas questões não é recente na filosofia, fazendo parte dos discursos epistemológicos ao longo da história. Mas, apenas nas últimas décadas a questão da revisão de crenças passou a ser abordada formalmente, em

[5] Podemos definir não-monotonicidade temporal da seguinte forma: Seja K um conjunto de crenças, A e B para subconjuntos de K, C(X) para a operação de fechamento dedutivo e K_t para determinado conjunto de crenças no tempo t, temos: $t_1 \leq t_2 \Rightarrow A_{t_1} \subseteq A_{t_2}$ e a não monotonicidade lógica como $A \subseteq B \not\Rightarrow C(A) \subseteq C(B)$. (NEBEL, 1992. p.71)

especial, com o desenvolvimento de modelagens de sistemas de inteligência artificial. O início deste desenvolvimento produziu frutos não apenas computacionais, mas filosóficos. Pelo caminho filosófico, passou-se a discutir as operações implícitas em nossos processos de aquisição e revisão do conhecimento e um sistema verossímil de representação de alocação destas crenças[6].

De modo mais pontual, a questão da revisão de crença trata situações em que nos deparamos com crenças inconsistentes. Há diversas situações em que a resolução dessa inconsistência é urgente, seja no âmbito pessoal ou coletivo. Ou mesmo em situações em que máquinas recebem *inputs* contraditórios, como por exemplo: Toda placa de "pare" no Brasil é vermelha. Pela configuração do GPS do meu carro inteligente estou no Brasil; Placas de "pare" são imóveis; em situações em que detectam-se placas de "pare", o sistema deve reduzir proporcionalmente à velocidade que estou e à distância em que a placa está de mim; O carro se desloca à 60km/h; Foi detectada uma placa de "pare" à frente; Mas, a placa se distancia do sensor em linha reta à velocidade de 10 km/h.

Nesta situação, a inconsistência por si, talvez não seja o problema para nós que podemos imaginar situações que descrevam este caso. Entretanto, imaginar isto envolve, em grande medida, introduzir novas regras ou eventos que não estão dados (talvez, para a qual a IA não tenha sido treinada). Além disso, escolher qual destes eventos e regras tem precedência sobre os outros. E como podemos ver intuitivamente, este modo de formalizar as crenças cai sob o conceito de lógica não-monotônica[7]. Especificamente, naquelas que tratam de inferências na presença de informação inconsistente (BREWKA. 1991. p.10).

O problema, de fato, é determinar qual das premissas (regras ou informações) devem ser descartadas para recuperar a consistência, uma vez que, a escolha do descarte pode modificar radicalmente as consequências que podem ser tiradas das crenças que permaneceram. A preocupação com as condições de manipulação destas crenças e o resultado dessa manipulação constitui o

[6] Pelo lado da computação, a possibilidade de formalizar estas operações garantiu grande agilidade no trato de bancos de dados e na programação de sistemas inteligentes.

[7] Definimos monotonicidade como $Se\ \Gamma \vdash A\ então\ \Gamma \cup \{B\} \vdash A$.

cerne da revisão de crença (*belief revision*). Mais precisamente, a revisão das crenças em função da chegada e uma informação nova que pode ou não conflitar com o conjunto de crenças já estabelecido (cf. DELGRANDE; WASSERMANN, 2013.).

3.1 O paradigma AGM

Dentre os modos de abordar esta questão, o ponto de partida será é o paradigma AGM (cf. ALCOURRÓN, GÄRDENFORS, MAKINSON, 1985; GÄRDENFORS, 1988), que visa o tratamento dinâmico de conjuntos de crenças[8]. O vocabulário básico deste paradigma é o seguinte (cf. GÄRDENFORS, 1988.):

Seja L uma linguagem lógica fechada sob os conectivos lógicos clássicos e orientada pela relação de consequência $\vdash$, satisfazendo as seguintes condições: (i) Se A é uma tautologia vero-funcional, então $\vdash A$; (ii) Modus Ponens: $\vdash A \rightarrow B$, $\vdash A$ então $\vdash B$; (iii) $\vdash$ é consistente (isto é, não é o caso que $\vdash \perp$.[9]

Definimos um conjunto Kde sentenças como um conjunto de crenças se e somente se (i) $\perp$ não é uma consequência lógica das sentenças em K e (ii) se $K \vdash B$, então $B \in K$. Note que esta *definição* evita que os conjuntos de crenças conduzam a contradições lógicas internas, esta pré-condição é chamada *integrity constraint* (GÄRDENFORS, 1992. p.6), por outro lado, implica em uma *onisciência lógica*. Considerando a lógica clássica, os casos em que K é inconsistente (temporariamente) dizemos: $K \vdash \varphi$, *para todo* $\varphi \in L$, expresso por $K_{\perp}$.

Dado um conjunto de crenças $K \in L$. Denotamos o conjunto de todas as crenças de L por K_L. K é completa se e somente se para qualquer crença $\varphi \in L$, $\varphi \in K$ ou $\neg \varphi \in K$. A totalidade das teorias consistentes e completas de L é expresso por M_L. Para todo conjunto de sentenças K de L, $[K]$ indica o

[8] Também chamada de *logic of theory change* e *theory of epistemic change* (NEBEL,1992. p.52).

[9] $\vdash$contém a lógica proposicional clássica, satisfaz o teorema da dedução ($\vdash A \rightarrow B$ *se e somente se* $A \vdash B$) e o teorema da compacidade (uma lógica é compacta sss para qualquer A, se A é uma consequência lógica de um conjunto de sentenças Γ, então há um subconjunto finito Γ'de Γ, tal que, $\Gamma' \vdash A$ (GÄRDENFORS, 1988. p.24).

conjunto das teorias consistentes e completas contendo K, seja K um singleto ou um conjunto maior. Para K e um conjunto Λ de sentenças de L, $K+\Lambda$ é fechado sob $\vdash$, tal que $Cn(K \cup \Lambda)$, como veremos a seguir. Note que, por se tratar de uma formalização epistêmica, o conjunto de crenças K não precisa ser maximal, pois seria interpretado como a onisciência factual do portador do conjunto, o que pode ser logicamente expediente, mas contraproducente no âmbito da aplicação.

O paradigma AGM clássico considera as seguintes operações básicas de mudança de crença (*belief change operations*) sobre K: (E_B) Expansão de crença, (R_B) Revisão de crença e (C_B) contração de Crença[10]. Dentre eles, nos ateremos à (R_B) e (E_B)[11]. Em razão da equivalência desenvolvida por Harper e a equivalência desenvolvida por Levi (*Ibidem*. pp.68-71.), (R_B) poder ser definida por (C_B) e vice versa.

Expansão de crença (E_B): A operação de expansão de crença ocorre quando realizamos a aquisição (a aprendizagem) de uma nova crença ao nosso conjunto, isto é, seja K consistente, em face de uma nova crença A onde $A \notin K \, e \, \neg A \notin K$, adicionamos A a K. Esta operação é denotada por K_A^+. Segue-se os postulados da E_B:

(K^+1) K_A^+ é um conjunto de crenças (*belief set*)

Este postulado garante que o input não descaracterizará o conjunto K e que K_A^+ é um par ordenado $K \times L$.

$(K^+2) A \in K_A^+$

Este postulado indica A como um *input* epistêmico válido.

$(K^+3) K \subseteq K_A^+$

[10] Para mais detalhes sobre a operação de contração (cf. DELGRANDE; WASSERMANN, 2013. p.488)

[11] As provas das propriedades das operações de expansão, contração e revisão estão em (GÄRDENFORS, 1988. pp.211-216).

Este postulado expressa o conceito de economia informacional em que, em vista de uma expansão, mantemos o máximo possível nossas crenças já adquiridas e evitamos $K_{\perp}$, entendido aqui como o super conjunto (*superset*) com todos os conjuntos de crenças ($K_{\perp}=\{\Gamma \subseteq L | P(\Gamma)\}$)[12]. Neste caso, isto significa impedir qualquer contradição com as crenças em K.

($K^{+}4$) Se $A \in K$, então $K^{+}_{A}=K$

Este postulado trata da aceitação de uma crença já contida no conjunto. Isto significa que reiterar a crença A não modifica em nada K.

($K^{+}5$) Se $K \subseteq H$, então $K^{+}_{A} \subseteq H^{+}_{A}$

O quinto postulado garante a monotonicidade lógica. No caso em que $B \notin K$, mas $B \in K^{+}_{A}$, temos que $B \in Cn(K \cup \{A\})$. Além disso, com os postulados de 1 a 5 podemos representar expansões conjuntivas como expansões reiteradas $(K^{+}_{A})^{+}_{B}=K^{+}_{A \wedge B}$, que vem a ter a mesma propriedade de comutatividade que a conjunção clássica $(K^{+}_{A})^{+}_{B}=(K^{+}_{B})^{+}_{A}$, logo $K^{+}_{A \wedge B}=K^{+}_{B \wedge A}$. Outra característica relevante para a exposição é o corolário (3.7) (GÄRDENFORS, 1988. p.51): Se $A \in K$, então $K^{+¿=K_{\perp}¿}_{A}$. Nas palavras do autor, a expansão não é capaz de revogar o estatuto de *inconsistência* de K (Para todo A, $(K_{\perp})^{+}_{A}=K_{\perp}$).

> ($K^{+}6$) Para todos os conjuntos de crenças K e todas as crenças[13] A, K^{+}_{A} é o menor conjunto de crenças que satisfaz ($K^{+}1$) - ($K^{+}5$).

Este postulado é econômico na medida em que garante a K^{+}_{A} que não existam novas crenças que não estejam vinculadas à A após a operação. O que em última instância significa dizer que $K^{+}_{A}=Cn(K \cup \{A\})$. Com estes postula-

[12] O que está implicitamente afirmado nesta convenção é que se K for ou se tornar inconsistente, *Ex Falso quod libet*, o que mais a frente chamaremos de *explosão*.

[13] No texto original A é uma sentença, mas manterei como crença para deixar que a roupagem externa (sentença, proposição, estado de coisas, informação, etc) possa ser escolhida de acordo com a teoria semântica de preferência.

dos podemos outras propriedades da expansão como a distributividade $(K \cap H)^{+}_{A} = \left(K^{+}_{A} \cap H^{+}_{A}\right)$ e a disjunção $K^{+}_{A \vee B} \subseteq K^{+}_{A}$ e $K^{+}_{A \vee B} \subseteq K^{+}_{B}$. Temos, então que $K^{+}_{A \vee \neg B} = K^{+}_{B \to A}$, logo, $K^{+}_{B \to A} \subseteq K^{+}_{A}$. Do mesmo modo, se $K^{+}_{A \vee A} \subseteq K^{+}_{A}$, então $K^{+}_{A \to A} \subseteq K^{+}_{A}$.

A revisão de crença K^{*}_{A}), por sua vez, é uma função que associa K e uma crença (ou conjunto delas) A, retornando um conjunto K' onde $A \in K'$. A diferença da revisão para a expansão é que no primeiro caso o conjunto K deve manter a *consistência* após a expansão sobre A. Para isto, K deve ter algumas crenças excluídas, o que significa que está operação é *não-monotônica*. Denota-se * com a operação de revisão (*AGM revision function)*. A operação de revisão, portanto, recebe como entrada K e $A \in L$ e retorna K' expresso como K^{*}_{A}. Os postulados da revisão são os seguintes (*Ibid.* p.54-56.):

(K*1) Para qualquer crença A e qualquer conjunto de crenças K, K^{*}_{A} é um conjunto de crenças.

Assim, (K*1) afirma que K não perde a propriedade de ser um conjunto de crenças sob a revisão.

(K*2) $A \in K^{¿}_{A}$

O que significa afirmar A e suas derivações a partir de K revisado.

(K*3) $K^{*}_{A} \subseteq K^{+}_{A}$

(K*4) Se $\neg A \notin K$, então $K^{+}_{A} \subseteq K^{*}_{A}$

(K*3) e (K*4) garantem a expansão de K por A: $(K \cup \{A\})$

(K*5) Se $K^{*}_{A} = K_{\perp}$ se e somente se $\vdash \neg A$

(K*5) assegura que após a operação, K^{*}_{A} permanece consistente.

(K*6) Se $\vdash A \leftrightarrow B$, *então* $K^{*}_{A} = K^{*}_{B}$

(K*6) afirma que revisão sobre sentenças sintaticamente equivalentes implica no mesmo conjunto de crenças final.

(K*7) $K^*_{A \wedge B} \subseteq \left(K^*_A\right)^+_B$

(K*8) Se $\neg B \notin K^*_A$ então, $\left(K^*_A\right)^+_B \subseteq K^*_{A \wedge B}$.

Estes dois últimos postulados governam a iteração de operações e a tradução de operações por conectivos e vice-versa.

Além destas operações, há diferentes critérios semânticos para ordem de prioridade das crenças que serão modificadas quando da aplicação das operações, como as funções de seleção (*selection functions*) que possuem a opção de modificar minimalmente K ou maximalmente (*maxichoice*); O enraizamento epistêmico (*epistemic entrenchments*) que organiza as sentenças do conjunto de crenças em função de sua resistência em ser revisada por estar relacionada com outras crenças, com vistas a uma nova sentença B, isto é, uma crença é mais útil e mais resistente à revisão, quanto mais ela estiver enraizada com outras. Ainda no mesmo mote, Nebel (Cf. 1989. pp.66-71.) fornece um critério de enraizamento, generalizando o enraizamento epistêmico com base em condições puramente sintáticas, chamado de relevância epistêmica (*epistemic relevance*). Há também o critério de probabilidade baseado no teorema de Bayes que calcula e seleciona a probabilidade de uma conclusão inferida com base na probabilidade de suas hipóteses; alternativamente, há o sistema de esferas, baseado em uma semântica modal, que atribui uma ordem *default* das teorias consistentes e completas do conjunto de crenças. (Cf. PEPPAS, 2008. pp.323-329);

Dentre estas opções, nos voltaremos ao enraizamento epistêmico para fundamentar um dos modelos de ordenamento usados por Galliers. Este critério de ordenamento lida com sentenças individuais no interior de um conjunto de crenças. Isto é, esta função ordena e decide o que serão feitas das sentenças quando o conjunto ao qual elas pertencem sofrer uma contração ou revisão. Há duas ideias chave que governam esta função (GÄRDENFORS, 1988. p.87): 1) É possível determinar o relativo enraizamento epistêmico de sentenças em um conjunto de crenças independentemente das operações que ocorram com o conjunto. 2) Quando um conjunto de crenças sofre operações, as sentenças modificadas são aquelas com menor enraizamento epistêmico. Di-

zemos que uma crença é mais enraizada epistemicamente quando ela está mais envolvida com outras, no sentido de que considerar outra em substituição à primeira diminui a capacidade inferencial do conjunto[14]. Neste sentido, Gärdenfors afirma que sentenças com forma de lei (*lawlike*) possuem geralmente um enraizamento epistêmico maior que generalizações acidentais. Pois, o poder explanatório de sentenças com forma de lei possui um poder explanatório maior. Ora, a ideia aqui parece muito aproximada à proposta que fiz ao adicionar ao banco de dados de Araújo uma instância contendo as regras inferenciais. Regras estas que possuem a forma de leis. Leis, como sabemos, tem a forma condicional, a mesma forma que a regras materiais apresentadas por Brandom. Isto fica evidente ao considerarmos a aplicação de seu ordenamento sobre sistemas legais (GÄRDENFORS, pp.101-102). Embora haja similaridades, ele reconhece que o ordenamento mais adequado para este caso seria um enraizamento deôntico (*deontic entrenchment*) sobre os conjuntos de leis ou normas. Talvez, por isso, como veremos não apenas aqui, mas ao falarmos de ordenamento em também em Galliers, as sentenças na forma de regras, normas ou leis tenham resistência maior a revisão. Seja por seu vínculo explanatório, seja por sua forma de ser expressa logicamente. O que justificaria um procedimento próprio de operação sobre estas regras, tendo em vista a própria preservação de suas propriedades pragmáticas e semânticas. O que parece ter levado a introdução de cláusulas Horn como representante destas regras[15] tanto por Evans, quanto tomadas por mim e que devem ser apropriadamente investigadas em uma oportunidade futura.

Os postulados que governam este ordenamento qualitativo de enraizamento são os seguintes: Dado um conjunto de crenças K,

[14] É importante considerar esta função de ordenamento, pois muitas delas como *maxichoice* e *system of spheres* utilizam a semântica modal para suas definições. E o enraizamento epistêmico utiliza uma noção de peso inferencial. Mas, que não se deve confundir com peso probabilístico, usualmente atribuído à funções de ordenamento como a baseada no teorema bayesiano. (cf. LEVI, 1983. p.165)

[15] Gärdenfors parece associar ao critério de ordenamento do enraizamento epistêmico a noção de paradigma khuniano. Onde a mudança substancial neste ordenamento significa um forte indicio de revolução científica (ou mudança de paradigma) na terminologia de Kuhn. (GÄRDENFORS, 1988. p.88). Isto endossa de modo mais firme a ideia de que o ordenamento signifique uma certa constelação de regras inferenciais que governam a atividade dos indivíduos.

(Transitividade) Para qualquer A, B e C se $A \leq_{EE} B$ e $B \leq_{EE} C$, então $A \leq_{EE} C$

(Dominância) Pra quaisquer A e B, se $A \vdash B$, então $A \leq_{EE} B$. O que significa que ser logicamente mais forte implica ser menos epistemicamente enraizado. A justificativa é obvia, se mantivermos o critério de fechamento sobre $\vdash$. Remover B do conjunto exigiria remover A também. Logo, B é mais resistente à remoção que A. Já A poderia ser removido sem que B o fosse, pois poderíamos ter um C tal que $C \vdash B$ também. O que se tem em vista aqui é uma economia informacional (GÄRDENFORS, 1988. p.89).

(Conjuntividade) Para todo A e B em K, $A \leq_{EE} A \wedge B$ ou $B \leq_{EE} A \wedge B$.

A ideia é que fazer uma retração sobre $A \wedge B$ exige que se faça uma retração sobre A ou sobre B. Seguindo o mesmo argumento do postulado anterior. Disto temos como lema a (Conectividade) que diz que para duas quaisquer sentenças A e B, $A \leq_{EE} B$ ou $B \leq_{EE} A$. Expandindo a noção de ordenamento, é possível aplicá-la a toda linguagem L. De tal forma que as sentenças fora de K sejam minimamente enraizadas:

(Minimalidade) Quando $K \neq K_{\perp}$, $A \notin K$ sss $A \leq_{EE} B$ para todo $B \in K$.

Isto também significa que as sentenças logicamente válidas são as mais enraizadas, pois, por definição do fechamento sobre $\vdash$, não podem ser descartadas:

(Maximalidade $\vdash$) Se $B \leq_{EE} A$ para todo B, então $\vdash A$.

3.2. Assimilando o increased coherence model *de Galliers*

Para Galliers, o sistema de revisão de crenças deve ser pensando em uma interação entre agentes, de fato, há duas condições de revisão, um *input* em um contexto não-comunicativo, usualmente coberto pelo paradigma AGM e através da comunicação de outro agente, que tem como prerrogativa a intenção de causar uma modificação nos estados doxásticos do seu interlocutor. Isto implica na ampliação da compreensão de ambiente de interação do indivíduo para um ambiente carregado de linguagem e comunicação (tanto em um sentido direto, em um diálogo, por exemplo, como em um sentido indireto, em casos de leitura de um texto ou uma placa). Esta forma de compre-

ender o *setting* de interação dos indivíduos se enquadra melhor ao que parece ser a dinâmica do jogo de dar e pedir razões da perspectiva do participante do jogo.

É claro que este modo de produzir revisões de crença não fornece um controle estrito sobre a consistência das revisões, mas é possível estabelecer um critério qualitativo para determinar qual conjunto de crenças será preferível, uma vez dadas as diferentes formas de executar a revisão, mas em um sentido de orientação das crenças e no sentido de expressão delas (GALLIERS, p.226), embora não no sentido de ser algo sobre o qual o indivíduo é capaz de raciocinar e computar sobre, como na modelagem autoepistêmica (BREWKA, 1991. pp.23-29). Nestes termos, a pergunta deve ser "qual é o estado mais coerente, dado meu estado cognitivo atual e esta mudança no ambiente que acabou de ocorrer?" (GALLIERS, p.222). Embora ela afirme que não há regras que conduzam a aceitação ou recusa da performance comunicativa por parte do indivíduo, a ação racional comunicativa deve ser pensada não apenas como proposital no sentido de conferir ou obter um estado doxástico, mas estratégica, no sentido de dividir o reconhecimento e controle dos efeitos da comunicação entre os participantes da situação, mediante princípios racionais compartilhados entre os indivíduos (cf. GALLIERS, 1989; 1991), tal como Brandom ao escolher a prática de contar escores de Lewis para formalizar a situação do jogo de dar e pedir razões, em que as próprias condições de condução coletiva da comunicação devam ser ditadas pelas regras materiais inferenciais e de incompatibilidade. Deste modo, sendo possível associar este sistema de regras materiais aos princípios racionais de Galliers.

Da perspectiva de uma expressão formal deste cenário, as condições de aceitação podem ser resolvidas com a introdução de funções e hierarquias no âmbito da atribuição de escores entre os falantes para a aceitação de estímulos não-inferenciais e asserções de terceiros a partir de estímulos não-inferenciais, como autoridade, confiabilidade, etc. Ferramentas estas, que podem ser introduzidas a partir da formalização de Evans. Já para a aceitação de asserções baseadas em inferências, podemos introduzir a atividade de explicitar e revisar as regras materiais que conduzem o contexto de comunicação, se-

guindo a afirmação de que os indivíduos não sabem de tudo sobre seu ambiente e que nenhum deles sabe sobre os estados doxásticos do outro.

Este cenário de ajuste constante entre evidências e argumentos enviados ao indivíduo e seus ajustes internos com vistas à coerência máxima entre seus estados doxásticos, Galliers chama de *cooperação adaptativa em um ambiente distribuído* (1992. p.223). Neste sentido, o ambiente é dinâmico e contém informações incompletas, além disso, não há axiomas governando a interação entre os agentes, apenas heurísticas que promovem a possibilidade de cooperação, que emerge dos princípios gerais de revisão de crenças e a autonomia dos agentes.

Embora o modelo AGM clássico lide com isto por meio de enraizamento epistêmico, parecendo fornecer condições pragmáticas de escolha, há defeitos neste e em outros critérios de ajuste na base de crenças, como a inacessibilidade e a inflexibilidade no ordenamento. No caso específico do AGM clássico, o enraizamento epistêmico leva em conta o poder explicativo e o valor informacional, mas não considera o problema das múltiplas extensões. Assim, a introdução de uma nova evidência produz dois ou mais conjuntos de crenças com extensão equivalentes, mas incompatíveis entre si, onde o problema característico ao contexto comunicativo em que a informação esteja sempre sujeita a ser rejeitada como não-confiável, abre a possibilidade para outras alternativas não previstas pelas operações de revisão padrão.

Deste modo, deve existir um critério adicional de ordenação além da coerência que leve em consideração a introdução de regras gerais para identificação e marcação de crenças ou estímulos, quando vindos de fontes confiáveis e não-confiáveis, por exemplo. A proposta feita por Galliers é combinar três teorias de mudança de crença racional: a teoria coerencial de crenças, a teoria justificacional de crenças, formando o sistema ICM (*increased coherence model*) e um conjunto de regras heurísticas. A primeira teoria é orientada pelo principio da conservação: “crenças correntemente e completamente aceitas estão justificadas na ausência de qualquer conflito com elas” (GALLIERS, p.230). Caso ocorra o conflito, segue-se o princípio da mudança minimal: na presença de uma revisão, devem-se fazer mudanças mínimas, tanto no acrés-

cimo de novas crenças, quanto na eliminação de antigas. E, por fim, o princípio de incremento de coerência: "as mudanças são permitidas apenas na medida em que elas trazem incremento suficiente na coerência" (Harman, 1986. Apud *Ibid.* p.230).

A segunda teoria se baseia no principio de justificação regressivamente até a parada em uma crença básica, considerada *autoevidente* no contexto comunicativo. O enraizamento epistêmico é um exemplo desta teoria. Uma crítica apontada por Galliers (*Ibid.* p.231) ao enraizamento é a dificuldade em representar as justificações das crenças e a propagação da descrença. Pois, há argumentos que mostram que os indivíduos não mantêm rastreadas todas as suas crenças, de modo que crenças não justificadas são mantidas e que a tentativa de manter o registro de todas as justificações excede a capacidade computacional, exigindo um custo alto demais. Suponho como alternativa para esta critica, em conformidade com Brandom, o estabelecimento de que as justificações são lances públicos. Portanto, o processo de *keep tracking* das crenças não é feito individualmente (mas, contando com um processamento distribuído), que também não é feito ilimitadamente e, para isto, estabelecemos como parâmetros de justificativa o contexto de comunicação como a Galliers faz, em adição ao já elaborado por Evans. Isto é, até a satisfação dos interlocutores, em casos mais extremos, na passagem da justificativa da crença para a justificativa da regra, levando, portanto, o processo a um patamar público de avaliação das regras em jogo, não mais como um processo de explicitação das crenças pressupostas. Neste sentido, e que parece estar de acordo com o que expusemos a partir de Brandom, Galliers apresenta o princípio do comprometimento positivo (*principle of positive undermining*): "Apenas pare de acreditar em uma crença ocorrente, se há razões positivas para fazê-lo, e isto não inclui a ausência de justificação para aquela crença". Isto é, uma crença só será abandonada se uma situação ou interlocutor problematizar a crença por meio de uma oposição, i.e. uma crença contrária ou uma solicitação de justificativa posteriormente comprovada inconsistente.

Assim, o ICM reage a uma nova evidência incorporando-a ao conjunto de crenças resultantes das alternativas logicamente equivalentes mais persis-

tentes, incluindo a possibilidade da nova evidência ser rejeitada. Isto é, o conjunto resultante é o mais difícil de ser desacreditado. (Ibid., 1992. p.232) Para fazer isto, ela introduz o critério *mc*, que similarmente ao enraizamento epistêmico, cria um ordenamento de justificação. Mas, de conjuntos consistentes de crenças e não de crenças individuais no interior de um conjunto de crenças. De tal sorte, ser permissível revisões alternativas de ordem sobre um mesmo ordenamento de conjuntos. A partir disto, ela apresenta dois postulados de operações sobre conjuntos de crenças.

1) A operação adição: $(K^{+}A)=\{K' \mid K' \subseteq K \cup \{A\}\ e\ cons(K')\ e\ K' \vdash A\} \cup \{K\}$

Onde, cons(K) significa K é consistente e pode ser definido como "existe um B tal que $K \nvdash B$". Note que K deve ser consistente, mas não deve ser fechado sobre dedução, diferente das condições AGM clássicas. Isto em conjunto com a definição, permite afirmar que nem sempre $K' \vdash A$. Uma vez que, deve estar sempre disponível a opção de rejeitar a nova sentença e, adicionalmente, $K^{+}A$ não produz necessariamente subconjuntos maximais de K, na medida em que, a meta *não* é manter o máximo de nossas crenças antigas com a chegada da nova sentença, afinal, podemos nos imaginar em situações em que estamos completamente enganados sobre o que ocorre.

2) A operação de deleção:

$(K \div A)=\{K' \mid K' \subseteq K \cup A\ e\ cons(K')\ e\ K' \nvdash A\} \cup \{K\}$

Esta operação estabelece apenas um parâmetro para a criação de novos conjuntos de crenças. A escolha de ordenamento dependerá da relação de ordenamento em questão, tal que, haja ao final um membro maximal que será o resultado da operação de revisão.

A relação escolhida combina um critério fundacional que se apoia na noção de conjuntos bem fundamentados. Um conjunto bem fundamentado são os membros conjunto teoréticos maximais de ¿. Assumindo-se um conjunto de crenças autoevidentes. Assim, para dois conjuntos K e K',

$K' <_J K$ *se para todo* $A \in K$, $K \cap E \vdash A$ e há um $B \in K'$, tal que, $K' \cap E \nvdash B$. A ideia é que K é mais suportada que K'. Este critério é usado em conjunto com outros três critérios coerenciais: a coerência máxima, a de menor mudança no conjunto de crenças e o critério *mc*. Os dois critérios anteriores já foram expostos. Restando *mc*, que é uma função ternária responsável por ordenar K a partir do aumento de sua capacidade de derivabilidade. Isto é, dadas as crenças fundacionais em dado contexto, é escolhido o conjunto de crenças que com o acréscimo de uma sentença nova, a confiança do indivíduo nas crenças centrais aumenta. Uma sentença A aumenta a coerência de K dado algum conjunto de crenças centrais Φ se adicionando A à K, produz-se uma nova prova de Φ a partir daí. Descrevendo o conjunto resultante como mais justificado, contendo mais provas que endossam as crenças centrais. Tais crenças centrais, não entendidas em um sentido fundacionista filosófico, mas consideradas crenças centrais para o contexto de revisão do indivíduo.

O procedimento é o seguinte: 1) remove-se Φ de K, gerando várias alternativas de conjuntos de crenças. 2) adiciona-se A à cada conjunto resultante de (1). 3) avalia-se cada conjunto se Φ foi inferido em ao menos um destes conjuntos. Este procedimento é representado pela expressão: $mc(K, \Phi, A)$ *sss há um* $K' \in (K \dot{-} \Phi)^{+} A, K' \vdash \Phi$. Desta forma, produzindo a seguinte relação de ordem:

> $K' \leq_{\Phi} K$ *sss para todo* A, *se* $mc(K', \Phi, A)$, *então* $mc(K, \Phi, A)$.
> (GALLIERS, p.235)

Mas, apenas estes dois critérios não são capazes de resolver o problema das múltiplas extensões. Em uma situação em que aceitar A e rejeitar A possuem o mesmo peso, é necessário contar com um critério adicional de escolha. Tal critério é o da autoevidência. Crenças autoevidentes são conjuntos de sentenças que supomos serem o caso, no sentido em que não evocamos justificativas anteriores para sua posse, senão apontando para certas regras (GALLIERS, 1992. p.236). De uma perspectiva brandomiana, podemos chamá-las de crenças não-inferenciais (*non-inferential beliefs*) ou crenças assi-

miladas por confiabilidade. Neste sentido, tais crenças são endossadas pelas regras que orientam seu contexto de aquisição. Regras heurísticas que se traduzem em *regras materiais* e relações de incompatibilidade. Galliers fornece dois conjuntos de regras: 1) as de endosso e 2) as de tratamento de inconsistência. As primeiras são de classificação de confiança sobre a fonte das sentenças (primeira mão, segunda mão, confiável, espúrio); especificidade ou generalidade, de valoração positiva (e.g. boa) ou de valoração negativa (e.g. ruim); hipotética. Já as segundas hierarquizam as marcações anteriores: crenças obtidas em primeira mão são mais resistentes à revisão que as de segunda mão; quanto mais crenças positivas comunicadas (sobre as negativas) ou especificidades assumidas (sobre as generalidades) que fundamentam um estado de crença (*belief state*), mais resistente ele deve ser; as valorações só são relevantes quando associadas às crenças em que negações em primeira mão são preferíveis que negações de segunda mão ou generalizações. Note que, entre as crenças centrais, Galliers admite que regras (materiais, da forma $(\alpha \wedge \beta) \rightarrow \delta$) podem ser endossadas ou abandonadas em função de novas sentenças (*Ibid.* p.235), nos colocando em contato direto com a ideia de revisão sobre regras. Afinal, a própria escolha da heurística que compõem a terceira camada de seu esquema de ordenamento de conjuntos de crenças pode ser colocada sobre questionamento: seria esta a única forma de construir hierarquias qualitativas? Ora, o conflito entre regras evoca a própria necessidade de possuirmos um aparato para revisá-las.

4. Conclusão

A ideia deste estudo é, antes de tudo, procurar a melhor forma de expressar logicamente a dinâmica complexa do jogo de dar e pedir razões. Operações internas, revisões das próprias regras, processamento distribuído de performances e não-monotonicidade são algumas das preocupações em vista. Mas, ainda parecem longe de alcançar a complexidade de nossos processos de compreensão, comunicação, organização e racionalização. Entretanto, parece servir como como resultado parcial de investigação sobre como modelar o complexo ambiente pragmático-inferencial descrito por Brandom. Este acrés-

cimo de Galliers parece nos deixar um pouco mais distante do que se havia chegado até agora. A sêxtupla combinada de Evans e Araújo, $Entitled \wedge Committed_x < A, T, E, I, B, Op(B/T/E)$ é substituída por $Entitled \wedge Committed_x < A, T, E, I, B, Op(K)$, sendo $Op(K) = (K^+ A) e (K \dot{-} A)$, de acordo com as três teorias combinadas por Galliers. Note que nas teorias dela, especialmente a heurística, nos coloca mais perto da relação entre nossos estados doxásticos e as regras de inferência material em conformidade com Brandom. Uma vez que, nos jogos de linguagem, para ele e, antes, para Lewis, devemos levar em conta as seguintes condições cinemáticas do *score-keeping practice*:

i. Cada lance abre novos cursos de ação e impossibilita outros. Esta impossibilidade e abertura são instituídas
ii. A correção de um lance é determinada em função do estado atual do jogo.
iii. As regras especificam a cinemática do jogo. Com isto, determinam as autorizações, compromissos e incompatibilidade de lances. Isto é, as circunstâncias de realização de determinados lances, suas consequências e os impedimentos. (cf. BRANDOM, 1998. p.183).

"Se em um tempo t o escore conversacional é s e se entre t e t' o curso da conversação é c, então no tempo t' o escore é s', onde s' é determinado de um certo modo por s e c." (Lewis, 2004. p.345)

iv. Os indivíduos, explicita ou tacitamente, concordam com as regras em exercício, mesmo as que conduzem a contradições, que porventura serão sanadas.
v. O histórico da comunicação e as regras do jogo constroem a marcação dos escores e a passagem de uma atribuição de escores à outra pode

ser expresso por uma função de transição entre uplas de ações, regras, histórico e participantes, representando etapas da comunicação.[16]

vi. É possível incorporar eventos externos ao jogo em curso. Isto é, eventos que não tenham sido trazidos na performance linguística dos indivíduos. Se o evento for aceitável[17], ele pode ser incorporado (LEWIS, p.347).

vii. No caso de ações incorretas (mas, não vazias), evoca-se a justificação da ação e, a partir daí, os interlocutores podem exigir retratação ou penalizações ao falante.

E, com isto, a modelagem de Evans acrescenta ao cenário o estado de composição entre regras e indivíduos fornecendo o seguinte algoritmo (HECK, 2018. p.103):

1. Iniciar assumindo que todas as asserções estão autorizadas, salvo casos explicitamente contrários.
2. Acompanhar o registro dos compromissos proferidos.
3. Computar as consequências das afirmações correntes autorizadas.
4. Remover todas as afirmações incompatíveis com compromissos já assumidos.
5. Repetir 2-4 até que a situação se resolva i.e. nenhuma performance não-vazia seja adicionada (*condição de parada*).

Da perspectiva do falante, o algoritmo filtra o que é considerável como *input* e o que não o é. Uma vez, reconhecido o *input*, iniciamos a modelagem da Galliers. Caso as performances não sejam reconhecidas, dizemos que são lances no vazio[18].

Se por um lado, são computados publicamente os escores, simultaneamente da perspectiva do indivíduo, um conjunto de operações estão em anda-

[16] Há uma sexta condição, de incorporação de eventos externos ao jogo. Não consideraremos esta característica

[17] Ser aceitável aqui é ter relevância como argumento entre os jogadores e estar de acordo com alguma das regras do jogo.

[18] Lances no vazio podem ser pensados como ações linguísticas que não se enquadram em nenhuma das regras do jogo, ou que não tem relevância para os agentes.

mento. Dado um conjunto inicial de crenças e um conjunto de endossos sobre estas crenças, o curso dos lances levará o indivíduo a 1) expandir as crenças, 2) revisar as crenças ou 3) ignorá-las. A escolha dependerá dos conjuntos disponíveis após o ordenamento de *mc* e a comparação dos endossos sobre cada conjunto de crenças disponível (cf. GALLIERS, 1992. pp.238-242.). Deste modo, a escolha do novo K determinará o escopo de ações do indivíduo diante dos lances de seus interlocutores, e a reação dele por causa disto.

É claro que ainda há elementos que necessitam de reformulação e eles parecem residir nas operações sobre cláusulas Horn. De fato, parece que elas habitam o conjunto de crenças do método de Galliers, assim como habitam o paradigma AGM clássico, de tal forma que, poderíamos estabelecer regras (*default*) para subscrevê-las nos conjuntos de crenças como uma classe especial de tratamento. Além disso, parece adequado estabelecer critérios materiais para avaliar a extensão de $Cn(K)$. Afinal, Brandom defende que a lógica não é a responsável por oferecer as regras de inferência do jogo de dar e pedir razões. Ela é apenas a nossa melhor ferramenta para expressar estas regras (é o que ele chama de expressivismo lógico). Embora Galliers tenha se posicionado pelo não fechamento sobre $\vdash$, as regras de inferência clássicas continuam subentendidas como válidas na sua formalização. Assim, acredito que seria interessante substituir as regras lógicas clássicas por um sistema de cálculo de sequentes, onde as regras das premissas à conclusão fossem definidas como pensou Brandom: condições de assertibilidade e suas consequências. Desta maneira, resgatando parte do holismo semântico defendido por ele na forma de subconjuntos de K, delimitados por fechamentos sobre regras de inferência material. Por exemplo, em uma situação como as regras para conjunção no sistema de Gentzen:

$$R\wedge: \frac{\Gamma, A, B \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta}$$

$$L\wedge: \frac{\Gamma \vdash \Delta, A \; \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \wedge B}$$

Γ passaria a significar $Cn(K)$, fechado, não sobre $\vdash$ em um sentido lógico, mas sobre um conjunto ou subconjunto de regras de inferência material, passando a ser representado pelo sinal $\vdash$ e o travessão de inferência significaria a passagem das condições de assertibilidade para a asserção (no caso da autorização) ou da asserção para as asserções possíveis (no caso do compromisso). Assim, poderíamos acrescentar à $\vdash$ um indexador M indicando inferência material ($\vdash_M$) e pensar estas inferências em uma classe $\{\vdash_{M_1}, \ldots, \vdash_{M_n}\}$, tal que, $\vdash_{M_n}$seja fechado sob a totalidade de regras reconhecidas pelo indivíduo, onde as relações de incompatibilidade poderiam ser introduzidas como regras *default*, tendo como consequência, também o fato de que $K_\perp$ passaria a ser entendido materialmente, tornando a inconsistência lógica, um caso particular de inconsistência. E, a partir daí, construir um *deotic entrenchment* das regras e estender este ordenamento para as operações sobre as cláusulas Horn contidas na sêxtupla. Neste sentido, a investigação tem as funções condutivas do debate D_X preservadas e o processo simplificado de Araújo substituído por um sistema autônomo de revisão construído para o contexto de comunicação, com a devida abertura para estender o procedimento de revisão para as próprias regras do sistema.

Podemos apontar ainda outra questão que merece investigações futuras: O tratamento para situações de crenças em conjuntos de crenças (e de ocorrência factual de regras) onde haja a persistência de inconsistências. A opção para o tratamento deste problema, em situações não solucionáveis por regras heurísticas de abandono da crença, como proposto por Galliers, talvez seja a adição de um filtro paraconsistente que sirva de *constraint* para os processos inferenciais com resultados fundamentalmente inconsistentes, evitando a parada do sistema (do jogo) e, igualmente, sua trivialização (que considerando o paradoxo da implicação material, está mais presente do que parece), que leva ao *ex falso quodlibet*. Galliers oferece critérios para acomodar e ajustar as crenças inconsistentes baseado nos critérios de coerência, justificativa e evita a trivialização, abdicando o fechamento sobre $\vdash$. Mas, se a inconsistência for persistente na medida em que os ajustes feitos pelo indivíduo sempre o leva a inconsistências? Considero este evento como paradigmático para a importân-

cia de operações de revisão, não apenas nos compromissos doxásticos dos indivíduos, mas podendo ser estendido às regras que orientam as inferências no sistema, isto é, nas regras inferenciais que levam o indivíduo a comprometer-se com crenças persistentemente inconsistentes. Esta forma de expressar logicamente o jogo de dar e pedir razões parece estar em consonância não apenas com os argumentos brandomianos trazidos à luz, mas com nossas intuições mais básicas do que significa navegar, se orientar e se organizar no mundo em que vivemos, um mundo que parece ser bem recheado de inconsistências.

Referências

ALCHOURRÓN, C. E.; GÄRDENFORS, P.; MAKINSON, D. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions". In: *The Journal of Symbolic Logic*, Vol. 50, No. 2. New York: Association for Symbolic Logic, 1985. pp. 510-530.

ARAÚJO, A. B. de. "Semantic Information and Artificial Intelligence". In: MÜLLER, Vincent C. (Ed.). *Fundamental Issues of Artificial Intelligence*. Oxford: Springer, 2016. p.129-140.

BÉZIAU, J.-Y.; CARNIELLI, W.; GABBAY, D. (Eds) *Handbook of Paraconsistency*. London: College Publications, 2007.

BRANDOM, Robert B. *Articulating Reasons: An Introduction to Inferencialism*. Cambridge: Harvard University Press, 2000.

_____. *Making it Explicit: Reasoning, Representing and Discursive Commitment*. Cambridge: Harvard University Press, 1998.

BREWKA, G. *Nonmonotonic Reasoning: Logical Foundation of Commomsense*. Cambridge: Cambridge University Press, 1991.

DARWICHE, A.; PEARL, J. "On the logic of iterated belief revision". In: *Artificial Intelligence*. Volume 89, No.1-2, Amsterdam: Elsevier, 1997. pp.1-29.

DELGRANDE, J.P.; PEPPAS, P. "Revising Horn Theories". In: Proceedings *of the Twenty-Second International Joint Conference on Artificial Intelligence*. Barcelona: AAAI Press, 2011. pp.839-844.

_____. "Belief revision in Horn theories". In: *Artificial Intelligence*. No.218, Amsterdam: Elsevier, 2014. pp.1-22.

EVANS, R. P. "Computer Models of Constitutive Social Practice". In: MÜLLER, V. C. (Ed.) *Fundamental Issues of Artificial Intelligence*. Oxford: Springer, 2016. [Cap. 23] pp.391-411.

GALLIERS, J. R. "Autonomous Belief Revision and Comunication". In: *Belief Revision*. GÄRDENFORS, P. (Ed.) Cambridge: Cambridge University Press, 1992. pp.220-246.

GÄRDENFORS, P. *Knowledge in Flux*: *Modeling the Dynamics of Epistemic States*. Massachusetts: MIT Press, 1988.

GÄRDENFORS, P. "The Dynamics of Belief Systems: Foundations vs Coherence Theories". In: *Revue Internationale de Philosophie*, vol. 172. pp.24-46.

GÄRDENFORS, P. Belief Revision: "An Introduction". In: *Belief Revision*. GÄRDENFORS, P. (Ed.) Cambridge: Cambridge University Press, 1992. pp.1-28.

HECK, R. L. *Informação a Partir de Regras: Considerações para a Solução do Escândalo da Dedução pela Semântica Inferencial Pragmática*. Tese (Doutorado em Filosofia) – Programa de Pós-Graduação em Filosofia. Universidade Federal do Ceará, Fortaleza, 2018.

HUHNS, M. N; SINGH, M. P. "Cognitive Agents". In: *IEEE Internet Computing*, 1998. pp.87-89.

LEWIS, D. "Scorekeeping in a Language Game". In: *Journal of Philosophical Logic* [1979] Vol. 8, No.3, Dordrecht: Kluwer Academic Pub., 2004. pp.339-359.

NEBEL, B. "Syntax Based Approaches to belief revision". In: *Belief Revision*. GÄRDENFORS, P. (Ed.) Cambridge: Cambridge University Press, 1992. pp.52-88.

PEPPAS, P. "Belief revision". In: VAN HARMELEN, F.; VLADIMIR, L.; PORTER, B. *Handbook of Knowledge Representation*, Netherlands: Elsevier, 2008. pp.317–359.

SELLARS, W. "Inference and Meaning". In: *Mind*. Oxford: Oxford University Press,1953. Vol. 62, No. 247. pp. 313-338.

_____. "Some Reflections on Language Games". In: *Philosophy of Science*. Vol. 21, No. 3. Chicago: University of Chicago Press, 1954. pp. 204-228.

WITTGENSTEIN, L. *Philosophical Investigations* [1953] (4th ed.). Oxford: Blackwell, 2009.

ANTI-EXCEPCIONALISMO E A METODOLOGIA DA LÓGICA

Evelyn Erickson[1]

RESUMO

Uma tendência recente na filosofia da lógica, sob o título de "anti-excepcionalismo", propõe que os princípios lógicos não são justificados por intuições *a priori*, mas ao invés, assim como nas outras ciências, por evidências *a posteriori* ou empíricas. O presente artigo apresenta o anti-excepcionalismo lógico atribuído a Quine, assim como duas propostas mais recentes para o método de revisão de teorias lógicas, a abdução e o equilíbrio reflexivo. Por fim, aborda-se o problema da lógica no pano-de-fundo, o qual põe em xeque a possibilidade de haver um método de revisão anti-excepcional para teorias lógicas, haja vista que qualquer método depende de alguma lógica para seu processo de revisão.

Palavras-chave: Anti-excepcionalismo lógico. Abdutivismo lógico. Equilibrio reflexivo. Metodologia lógica.

ABSTRACT

A recent trend in the philosophy of logic, which is usually called “anti-exceptionalism about logic”, proposes that logical principles are not justified by *a priori* intuitions, but rather, like in other sciences, by *a posteriori* or empirical evidence. The current essay presents the variety of anti-exceptionalism attributed to Quine, as well as two more recent proposals for the method of revision of logical theories, namely, abduction and reflexive equilibrium. Finally, the background logic problem is addressed, which puts into question the possibility of there being an anti-exceptional method of theory revision for logical theories, given that any such method will somehow depend on some logic for its revision process.

Keywords: Anti-exceptionalism about logic. Logical abductivism. Reflexive equilibrium. Methodology of logic.

1. Introdução

Em uma visão tradicional, o conhecimento lógico é fundacional e necessário, e sua evidência vem de intuições *a priori*. As ditas “leis lógicas” for-

[1] Doutoranda em Filosofia no PPGFIL (UFRN).
Email: eerickson@ufrn.edu.br. ORCID: 0000-0001-7710-8245.

mam a base do nosso conhecimento lógico, e antecedem qualquer experiência empírica. Tal tipo de conhecimento é distinto do conhecimento científico que, em contraste, se baseia em evidências empíricas e cujas teorias são revisáveis. A essa visão tradicional tem-se recentemente dado o rótulo de excepcionalismo lógico.

Um exemplo claro da história da lógica[2] é Gottlob Frege, que pressupõe "que o conhecimento lógico é privilegiado em seu status fundacional"[3] (Martin 2020, 4). Enquanto que Frege defende algo como a lógica clássica, um excepcionalista não precisa comprometer-se necessariamente ao monismo lógico ou à lógica clássica. Um exemplo de um excepcionalista pode ser encontrado também em Michael Dummett, que defende que a teoria lógica correta é a lógica intuicionista (Hjortland e Martin 2020).

Alternativamente, o anti-excepcionalismo lógico propõe que o conhecimento lógico é revisável, e não se baseia apenas em evidências *a priori*. Dessa forma, anti-excepcionalistas fornecem diferentes perspectivas sobre lógica e propõe distintos métodos de revisão.

Enquanto que divergências lógicas não são novidade na história da lógica, após o advento da algebrização da lógica e dos desenvolvimentos na filosofia da lógica no começo do século XX, a hegemonia de algo como a Lógica Clássica foi estabelecida. Nesse contexto, o excepcionalismo lógico fazia sentido, pois todas as evidências apontavam para a irrefutabilidade dos princípios clássicos, e assim tinha-se que a lógica clássica era a única lógica correta. Não obstante, muitas lógicas não-clássicas vieram a ser desenvolvidas e tal excepcionalismo começou a ser questionado. Aceitando-se intuições como evidência, e há lógicos com intuições divergentes, qual seria a lógica "correta"? Aceitando-se a analiticidade como evidência, já que isso é contingente na linguagem, qual seria a lógica "correta"? Nesse cenário, o anti-excepcionalismo lógico surge para prover uma nova epistemologia à lógica e fazer sentido dos

[2] É digno de nota que essa atribuição de "excepcionalista" aos autores mencionados é em certa parte revisionista, dado que eles não definiram suas próprias posições nesses termos, assim como alguns dos autores mencionados como anti-excepcionalistas também não usam esse rótulo.

[3] Todas as traduções aqui presentes são de própria autoria, exceto onde indicado na bibliografia.

desenvolvimentos das lógicas não-clássicas, assim como explicar a divergência de intuição entre os especialistas.

O presente artigo se estrutura da seguinte forma: a Seção 2 discute o surgimento do anti-excepcionalismo lógico, assim como dois exemplos presente na literatura, a saber, o abdutivismo lógico (Williamson 2016; Hjortland 2016, 2019; Priest 2006, 2016) e o equilíbrio reflexivo (Resnik 2004); Seção 3 aponta o problema da lógica no pano-de-fundo contra essas posições; por fim, na Seção 4 apresenta-se outros horizontes de pesquisa referente ao anti-excepcionalismo na lógica.

2. Anti-excepcionalismo lógico

W. V. O. Quine é apontado como o precursor do anti-excepcionalismo lógico,[4] em especial no seu influente "Two Dogmas of Empicirism", em que ele argumenta contra a distinção entre conhecimento analítico e sintético (em tal visão excepcional, a Lógica seria um tipo de conhecimento analítico). No mesmo ensaio, ele apresenta sua proposta de que o conhecimento se organiza em uma teia-de-crenças, e propõe uma visão holística sobre revisão dessas crenças. Para o holismo quineano,

> [a] totalidade de nossos assim chamados conhecimentos e crenças das mais casuais questões de Geografia e História até as mais profundas leis da Física atômica ou mesmo da Matemática pura e da Lógica, é tecido feito pelo homem, que encontra a experiência apenas nas extremidades. (Quine 2011, 66)

Ou melhor ainda, tal totalidade de conhecimentos é "um campo de força, cujas condições limítrofes são a experiência" (Quine 2011, 66). Nessa visão, a ciência (e o conhecimento em geral) depende tanto da linguagem quanto da experiência, e esses dois aspectos não podem ser tidos como separados: "a unidade de significância empírica é o todo da ciência" (Quine 2011, 66), aí estando a sua rejeição à distinção entre o conhecimento analítico e sintético.

[4] Quine não utiliza esse termo sobre sua própria filosofia. Williamson (2007) discute o "excepcionalismo filosófico", e passar a utilizar "anti-excepcionalismo sobre filosofia" (Williamson 2013), e só então "anti-excepcionalismo lógico" ou "anti-excepcionalismo sobre lógica" passa a ser usado.

Outro aspecto importante é como se dá a revisão desse campo quando há um novo *input* da experiência: "um conflito com a experiência na periferia ocasiona reajustes no interior do campo" (Quine 2011, 67). No holismo quineano, "nenhum enunciado está imune a revisão" (Quine 2011, 67), nem mesmo as chamadas leis lógicas (já que pode-se optar por uma revisão profunda —no cerne lógico—para acomodar o conflito na periferia).[5]

Vale também mencionar que a visão de Quine sobre o tópico do lugar da lógica na teia-de-crenças não é unívoco. Como argumenta Arnold e Shapiro (2007), existem dois Quines, um "amigo da lógica" e um radical. Enquanto que o Quine amigo da lógica aceita que as verdades lógicas são analíticas no sentido tradicional, o Quine radical inclui a lógica na mesma teia-de-crenças tornando-a não imune à revisão. É o Quine radical que, no entanto, exerce influência no desenvolvimento do anti-excepcionalismo lógico. Os autores mencionam também que para o Quine amigo da lógica, a lógica não está na teia, mas compõe a ligação entre os conhecimentos.

Dessa tradição quineana surge um anti-excepcionalismo lógico, que mantém que a lógica é um tipo de conhecimento também sujeito a revisão, e dessa forma não possui um caráter excepcional em relação às outras ciências. Uma teoria lógica responde ao mesmo tipo de evidência que inicia revisão de qualquer outra teoria. Embora os diferentes autores divirjam no que eles consideram ser uma teoria lógica, há alguns aspectos em comum. Primeiro, uma teoria lógica não é apenas um mero sistema formal ("lógica pura", segundo a nomenclatura de Priest (2006)), ou uma lógica aplicada a algum problema ("lógica aplicada"), e sim um sistema mais robusto para explicar a validade de argumentos ("aplicação canônica"). Segundo, a noção central da lógica é a validade, embora lide também com outras noções (como verdade, implicação, etc.). No geral, a validade é tida como um fenômeno independente da mente e

[5] Quine defendeu também a "máxima da mutilação mínima" como critério de revisão para teia-de-crenças, então não é claro até que ponto qualquer revisão central seria aceita (Hjortland 2017b).

da linguagem.[6] E terceiro, existe alguma justificativa para a revisão de teoria lógica, ou seja, um método que justifique essa revisão.

A partir do holismo de Quine, outras abordagens surgiram para explicar o caráter revisável das teorias lógicas. São aqui exploradas duas tais abordagens, elucidando os seguintes aspectos: o que é uma teoria lógica, qual evidência instiga uma revisão de teoria e qual o método dessa revisão. Jack Woods (2019) apresenta os dois métodos apontados aqui como comparação de teorias inteiras vs. abordagem fragmentada.[7] Enquanto que a comparação de teorias busca comparar teorias lógicas inteiras (como por exemplo a lógica clássica vs. a lógica intuicionista), a proposta fragmentada busca consertar problemas na teoria atual (qualquer que seja) aos poucos (por exemplo, descobrir como retirar a lei do terceiro excluído de uma teoria clássica, e quais modificações serão necessárias para tal, haja vista a necessidade de lidar com intuições sobre vagueza). Como exemplo do primeiro método, será discutido o método da abdução, e como exemplo do segundo método, será discutido o equilíbrio reflexivo, haja vista que são as propostas mais promissoras de cada tipo.

2.1. Abdução

Esta abordagem anti-excepcionalista propõe empregar o método de revisão chamado de abdução,[8] ou "inferência para a melhor explicação", tomado emprestado das ciências empíricas, na revisão de teorias lógicas. É digno de nota que o método abdutivo de revisão é comum na literatura do

[6] Tal independência se dá em oposição ao sentido carnapiano de dependência da linguagem, e não de forma absoluta. No sentido absoluto, há tanto anti-excepcionalistas que mantém a independência em relação a linguagem (como Williamson) como dependência (como Gillian Russell). Há também os que mantém que a lógica aceita evidências linguísticas (a ser tratado na próxima seção), o que não implica que a validade é um fenômeno linguístico. Nesse caso, evidências linguísticas trazem a tona certos aspectos subjacentes de um fenômeno não linguístico.

[7] Em inglês, *whole theory comparison* e *piecemeal approach*.

[8] Abdução é um modo de inferência proposto primeiramente por C. S. Peirce, que junto da dedução e indução são os tipos de inferência utilizados na investigação científica (Rodrigues 2011). Na literatura recente, no entanto, esse termo perdeu esse sentido original e veio a ser usado como método de justificação de seleção de teorias, sob o nome de "inferência para a melhor explicação"; ver Douven (2017).

realismo científico, mas é contestado na filosofia da ciência mais amplamente[9]. Tal método consiste em selecionar a teoria mais explicativa, o que é medido de acordo com algumas virtudes teóricas, tais quais simplicidade, força, adequação aos dados, etc. O realismo científico diz que a teoria mais explicativa é a que provavelmente é verdadeira.

Para os anti-excepcionalistas abdutivistas (a partir de agora, simplesmente abdutivistas), uma teoria lógica visa explicar o fenômeno da validade. Não se tenta explicar apenas "validade-em-L", isto é, validade de acordo com um sistema lógico L, mas sim validade *simpliciter*, cujas propriedades, acredita-se, podem ser descobertas e formalizadas. Uma teoria lógica é dessa forma semelhante a teorias científicas, já que essas também descrevem fenômenos independentes da mente e da linguagem. Como será mostrado, há ainda assim sutis diferenças nas abordagens dos abdutivistas no que diz respeito ao caráter ontológico da validade, já que esses autores aceitam diferentes tipos de evidência em seus argumentos abdutivos.

Sobre evidência lógica, Ole T. Hjortland (2019) propõe que há três tipos de evidência que podem ser considerados, em particular por abdutivistas. Primeiro, intuições (pré-teóricas) sobre validade (sendo assim uma fonte de evidência *a priori*[10]); segundo, compatibilidade com teorias não-lógicas (também chamado de argumento da incompatibilidade); e terceiro, evidências não-metalinguísticas.

Intuições (pré-teóricas) sobre validade constituem a evidência já utilizada por excepcionalistas. Por exemplo, uma intuição é que nossas práticas inferenciais não deveriam decidir questões de existência, e por isso "há pelo menos dois indivíduos" não deve ser uma lei lógica (Resnik 2004, 181), ou então a intuição de que é impossível um objeto ter e não ter certa propriedade ao mesmo tempo leva a "A ou não-A" ser considerada uma lei lógica. Dado um número suficiente de intuições, é possível compor uma teoria lógica mais

[9] Um ponto relevante é que os anti-excepcionalistas abdutivistas aceitam esse método por demais rapidamente, sem considerar os comprometimentos ontológicos que tal método trás para a lógica; ver Erickson (2020).

[10] É digno de nota que ao utilizar evidência excepcional e não empírica, o uso da abdução pode não mais ser uma metodologia justificada, do ponto de vista de utilizar o mesmo método das ciências.

robusta (não apenas uma lista de formas válidas de argumentos). Essa teoria lógica irá validar ou não outras formas de argumentos, como a lei de Peirce[11] ou outras formas que parecem contra-intuitivas. "Uma teoria lógica, então, é uma candidata a explicação de um conjunto de intuições sobre a validade de argumentos na linguagem natural" (Hjortland 2019, 259).

Tal processo de elaboração de teorias lógicas já é utilizada por excepcionalistas lógicos. Anti-excepcionalistas, no entanto, aceitam também outros tipos de evidência, se diferenciando assim dos excepcionalistas. Dessa forma, as teorias lógicas devem não só acomodar intuições (que são evidências *a priori* e não-empíricas), mas também algum dos outros dois tipos de evidência.

O segundo tipo de evidência abrange teorias não-lógicas, como teorias sobre vagueza, teorias sobre o predicado de verdade, e a aritmética, de modo que teorias lógicas devem ser compatíveis com tais teorias. Os paradoxos semânticos são de particular importância.

> [N]o caso dos paradoxos semânticos, o problema é não só que a lógica clássica produz consequências evidenciais indesejadas, mas que ela é simplesmente inconsistente com as afirmações não-lógicas. (Hjortland 2019, 268)

O exemplo mais comum do uso de compatibilidade entre teorias se dá quando uma mesma teoria baseada na lógica clássica tenta acomodar tanto os axiomas de Peano para a aritmética quanto o esquema irrestrito do predicado de verdade (V("A")↔A), numa tentativa de aplicar uma mesma teoria universalmente. Nesse âmbito, o fecho lógico dessa teoria é trivializado, dado que é classicamente inconsistente. A incompatibilidade entre essas três teorias é tida como evidência contra a lógica clássica.

> O classicista está livre para rejeitar o antecedente de teorias semânticas para salvar sua teoria lógica, mas isso vem com um preço. Da mesma forma, já que um não-classicista rejeita princípios clássicos, ele precisa providenciar uma alternativa para demonstrações clássicas. (Hjortland 2019, 269)

Dentre as diversas teorias que devem ser compatíveis com uma teoria lógica, muitos destacam teorias matemáticas, e o argumento mais comum é o argumento de indispensabilidade da lógica clássica para a matemática. Tal ar-

[11] $((\Phi \rightarrow \Psi \rightarrow \Phi) \rightarrow \Phi$

gumento é usualmente rejeitado, tanto porque questiona-se a matemática ser mais fundamental do que teorias da verdade, quanto a lógica clássica ser de fato indispensável, dado que lógicas não-clássicas conseguem recapturar[12] a classicalidade em casos restritos (Hjortland 2017b).

O terceiro tipo de evidência, evidências não meta-linguísticas, são proposições com conteúdo empírico, como por exemplo, uma teoria sobre a física (Γ), cujas proposições são sobre o mundo. Considerando $Cn_{\models}(\Gamma)$ o fecho lógico de Γ sob um operador de consequência $\models$, as sentenças de $Cn_{\models}(\Gamma)$ são também sentenças normais sobre o mundo, e não sobre validade ou outro conceito lógico. Tais teorias servem como evidência quando compara-se duas teorias lógicas, pois quando comparamos $\models_1$ e $\models_2$, "[o] resultado é confirmação ou desconfirmação indireta de teorias lógicas, sem evidência especificamente lógica" (Hjortland 2019, 264). A junção entre teorias lógicas e teorias científicas compõe uma visão mais naturalizada da lógica. Note que enquanto que o segundo tipo de evidência lida com teorias linguísticas (sobre verdade, vagueza, etc), as teorias desse terceiro tipo são teorias empíricas, sobre o mundo, e não sobre propriedades de sentenças.

Seguindo a tradição quineana, o abdutivismo lógico foi proposto sob esse nome primeiramente por Timothy Williamson (2007), que embora diga que "a avaliação de lógicas é contínua com a avaliação de teorias científicas, da mesma forma que Quine sugeriu" (Williamson 2017, 334), não segue o método de revisão proposto por Quine, e sim propõe o método abdutivo para selecionar a teoria lógica correta. Para Williamson, a melhor teoria lógica pode ser escolhida com base em uma "inferência para a melhor explicação". As virtudes teóricas que ele propõe são encaixe com a evidência (ou ao menos consistência com a evidência), força (dedutiva), simplicidade, elegância e poder unificador.

Evidências para Williamson são do segundo e terceiro tipo apresentado por Hjortland (2019): o argumento de indispensabilidade para a matemática e evidências não meta-linguísticas. Não surpreendentemente, o argumento

[12] Em inglês, *recapture*. A recaptura diz respeito a haver um método para que uma teoria lógica consiga expressar as mesmas inferências que outra teoria, mas em outro vocabulário. Para mais, ver Hjortland (2017b).

abdutivo de Williamson favorece a lógica clássica, já que é essa a lógica que é particularmente útil para a matemática. Devido ao alto custo de revisão, a lógica clássica não deve ser revisada como o cânone da inferência.

Outra proposta abdutiva é apresentada por Graham Priest (2016), que propõe dar diferentes pesos a cada virtude teórica, e propõe um cálculo simplificado para escolher entre teorias. Para ele, as virtudes relevantes para selecionar teorias lógicas são adequação aos dados, simplicidade, consistência, poder expressivo e evasão de elementos *ad hoc*.

Para Priest, a evidência lógica é principalmente do primeiro tipo apresentado por Hjortland (2019):

> É claro o suficiente o que provém os dados no caso das ciências empíricas: observação e experimento. O que desempenha esse papel na lógica? A resposta, eu creio, é as nossas intuições sobre a validade ou não de inferências vernaculares. (Priest, 2016, 41).

Além disso, inclui também o segundo tipo, já que ele aceita que "a teoria lógica é moldada por nossa confiança em outras afirmações, por exemplo, afirmações sobre nossas melhores teorias não-lógicas" (Hjortland 2019, 262). Em particular, os paradoxos semânticos são tidos por Priest como particularmente relevantes (Priest 2006).

Notoriamente, Williamson e Priest divergem em relação a qual evidência levar em consideração (quais dados devem ser explicados), e por isso divergem também em qual é a melhor teoria lógica. Hjortland (2017a, 2019) discute a divergência entre Priest e Williamson, apontando que ela se dá porque em seu argumento abdutivo, Williamson dá mais peso à força dedutiva de uma lógica, enquanto que Priest prefere preservar a compatibilidade com teorias de verdade tarskianas. Williamson utiliza um argumento de indispensabilidade da lógica para a matemática, mas Hjortland (2017a, 2019) contra-argumenta que teorias sobre verdade são tão fundamentais quanto, e além disso, há outras maneiras para lógicas não-clássicas serem fundamentais para a matemática, através de princípios de recaptura.

Por sua vez, Hjortland (2017a) propõe como evidência em seu argumento abdutivo além de intuições sobre a validade de inferências e compatibilidade com teorias e prática matemática, incluir também considerações "da

psicologia do raciocínio, normas epistêmicas e de racionalidade, e assim por diante' (Hjortland 2017a, 14).[13] Levando em conta essa ampla gama de evidências leva Hjortland (2017a) a concluir que a teoria lógica mais adequada é um tipo de pluralismo em que "validade não é uma propriedade monolítica" (Hjortland 2017a, 25). Isto é, dentro de uma teoria lógica, há diferentes princípios lógicos que podem valer para diferentes partes da linguagem. Ele chama tal pluralismo de *intra-teórico*, para contrastar com o pluralismo inter-teórico, no qual para cada noção de validade há uma teoria. Dessa forma, ele tenta acomodar o papel da lógica tanto nos fundamentos da matemática, quando nas teorias de verdade e em teorias de raciocínio.

Em suma, os abdutivistas lógicos concordam que a validade é um fenômeno independente da mente e linguagem[14] e que o método de revisão é "inferência para a melhor explicação", isto é, um argumento abdutivo, mas discordam em relação a qual evidencia uma teoria lógica deve explicar.

2.2. Equilíbrio reflexivo

O segundo método de revisão de teorias lógicas apresentado aqui é o "equilíbrio reflexivo", proposto recentemente por Michael Resnik (2004), mas já antecipado por Nelson Goodman (1955). O método foi assim denominado por John Rawls (1971) no contexto de revisão de teorias normativas. O método consiste em equilibrar "teoria" e "dados", isto é, equilibrar as inferências permitidas pelas teorias lógicas e os dados (evidências) que tenta-se acomodar.

> Aqui encontramos lógicos fazendo uso tanto de dados sobre nossas práticas inferenciais quanto de suas intuições—sendo estas tanto normativas quanto metafísicas—sobre os fatos da lógica. (Resnik 2004, 181)

A partir desses dados, um lógico produz uma teoria lógica, que consiste em "um sistema formal, uma semântica para esse sistema, uma metateoria

[13] Esse tipo de evidência pode parecer plausível, mas deve-se atentar para os perigos de se chegar a um psicologismo.

[14] Ver nota de rodapé 3.

que o atenda, e uma método de tradução para formalizar argumentos informais" (Resnik 2004, 180).

Tal proposta dificilmente será satisfatória de imediato, pois é possível que a teoria lógica inicialmente produzida proclame certos relacionamentos lógicos indesejados que de antemão não se previa, ou então é possível que o sistema proposto mesmo que não capture a intuição inicial, se mostre elegante demais para ser descartado. No primeiro caso será necessário adequar a teoria aos dados, e na segunda, é possível rejeitar os dados e preferir manter a teoria.

> O processo vem ao menos a um fim temporário quando os lógicos chegam em um estado de `equilíbrio reflexivo', isto é, quando eles consideram que sua teoria não rejeita supostos fatos lógicos que eles estão determinados a preservar e em contrapartida aceita nenhum que eles estão determinados a rejeitar. (Resnik 2004, 181)

O processo de revisão de lógica se dá então através de um vai-e-vem entre considerações teóricas e considerações evidenciais até que um equilíbrio seja atingido. A teoria lógica desenvolvida não será totalmente satisfatória nem para um lado nem para outro, de modo que caso haja novas evidências ou novas questões teóricas, o processo continua até que um novo equilíbrio seja encontrado.

A revisão de um sistema normativo se inicia a partir de um comportamento desviante por parte de alguns membros da comunidade, gerando novas intuições que por sua vez impulsionam o sistema a um equilíbrio reflexivo. Nessa visão, a lógica é anti-excepcional perante as disciplinas normativas.

Na lógica,

> [a]ssim como executamos essa prática na maioria dos casos sem pensar, nós também revisamos ela na maioria dos casos sem pensar. Nós simplesmente não mais aceitamos certas conexões inferenciais específicas, ou, mais frequentemente, nós reconhecemos novas conexões. Assim como mudanças numa língua, certas práticas inferenciais podem vagarosa e silenciosamente se tornar obsoletas e novas práticas podem evoluir sem cerimônia. (Resnik 2004, 189)

Uma mesma prática pode dividir-se em duas, como por exemplo, a chamada lógica clássica e a lógica intuicionista hoje em dia constituem duas práticas diferentes, mas ambas vêm de uma mesma linhagem. Nesse caso específico de revisão, a lógica intuicionista não veio a ser desenvolvida através

de uma mudança aleatória na prática inferencial, mas sim refletidamente a partir de uma certa intuição. Em geral, lógicos quando propõem mudanças não o fazem "a torto e a direito", mas apenas quando há uma proposta bastante razoável para a resolução de uma tensão na teoria vigente. No caso em questão, para alguns lógicos havia tensão demais entre a teoria da lógica clássica e a intuição (tida aqui como evidência) de que demonstrações deveriam ser construtivas.

Resnik, assim como os abdutivistas, discorre sobre o papel da lógica nas ciências empíricas. Ele mantém que a lógica é metodologicamente *a prioiri*, dado o papel que a lógica tem na formação de teorias na ciência. Está é uma marcante diferença entre os dois métodos aqui discutidos: enquanto proponentes do abdutivismo lógico consideram a lógica como uma teoria descritiva (já que ela mesmo seria uma ciência num sentido mais estrito do termo), Resnik considera a lógica como uma teoria normativa.[15]

> As leis da lógica estão entre nossos princípios metodológicos mais fundamentais. Em vista disso, tratarei a revisão da lógica como um caso especial de revisão de metodologia. Metodologia, tenho por mim, é um sistema de normas que governa nossa prática científica (e talvez nossa prática epistêmica de modo mais geral). (Resnik 2004, 184)

A noção de *a priori* empregada por Resnik é mais fraca do que a concepção tradicional de *a priori*, dado que considerações empíricas podem levar a revisão da metodologia vigente, e essa metodologia não está imune de revisão por razões não-empíricas. Dessa forma, Resnik mantém que

> [a]o invés de manter que nós adquirimos normas por via apriori, eu mantenho que nos encontramos com uma coleção de normas e valores culturalmente condicionadas, as quais podemos ou não modificar à luz da experiência, argumentos e mudanças em nossa condição. (Resnik 2004, 185)

Assim, a revisão de teorias lógicas se dá através de evidências empíricas à moda anti-excepcionalista. Embora Resnik não atribua tal rótulo para sua posição, é perfeitamente plausível que ele seja considerado um anti-excepcionalista.

[15] Mais sobre esse ponto é discutido na seção 3.2.

O método do equilíbrio reflexivo é atribuído também a Quine—tanto por Resnik (2004) quanto por Woods (2019)—no que diz respeito a mudanças na teia-de-crenças. É digno de nota que Resnik critica a abordagem quineana do método, dado que Quine tenta naturalizar a normatividade da lógica, o que ocasiona problemas.

Para Quine (o Quine radical, na terminologia de Arnold e Shapiro (2007)), a lógica é ao mesmo tempo o sistema de normas que governa a revisão na teia-de-crenças e a teoria que está no cerne da teia. Enquanto que a lógica como teoria está apta a revisão perante a experiência, o mesmo não é verdade para a lógica como o sistema de normas. Dado que Quine e outros naturalistas "não querem reconhecer algum tipo de base apriori para nossas crenças normativas ou nossas revisões delas" (Resnik 2004, 184), o que justifica então a escolha dessa lógica? Deve assim haver algum conhecimento *a priori* que guia a escolher de normas na teia-de-crenças. A abordagem de Quine tenta naturalizar a aprioricidade metodológica da lógica, mas não consegue fazer isso sem encontrar um sério problema.

Em contramão, Resnik adota uma abordagem não-cognitiva (ou não-factualista) sobre a normatividade, não sendo necessário naturalizar a base do conhecimento lógico, "[d]e modo que no caso em questão não há o que saber, e não há conhecimento epistêmico normativo para se naturalizar" (Resnkik 2004, 185). Rejeitando a abordagem quineana,[16] Resnik atribui a força normativa da lógica às práticas de uma comunidade.

> Se muitos membros de uma cultura, ou pelo menos membros influentes suficientes, se engajam em uma conduta inicialmente desviante, as normas da cultura podem mudar para legitimar essa conduta. Isso é o caso para metodologia, assim como para língua e moralidade. (Resnik 2004, 186)

Tal conduta desviante irá por vez servir como um novo dado no processo de equilíbrio reflexivo. Sendo assim, essa conduta irá ser rejeitada em prol da metodologia vigente, ou irá servir como ponto de partida para o desenvolvimento de uma nova metodologia. No âmbito da lógica, o reconheci-

[16] Enquanto que esse é um ponto em que Resnik discorda de Quine, é também um que os abdutivistas concordam com Quine. Para esses últimos, a força normativa da lógica vem de ser uma teoria "naturalizada" sobre fatos lógicos.

mento de padrões inferenciais fora dos já aceitos pela comunidade irá ocasionar uma rejeição da validade dessas inferências, ou então tais inferências serão aos poucos aceitas, modificando a própria prática (como por exemplo, tem acontecido com a rejeição da Lei do Terceiro Excluído e o desenvolvimento da lógica paraconsistente).

Uma preocupação (que será mais elaborada na seção 3.2) é que o ajuste de uma tensão entre teoria e dados não seja unívoco.

> [D]ado que pode não haver uma maneira única e ótima para responder ao problema intelectual, seja ele na ciência ou em outro lugar, é bem provável que a revisão de práticas inferenciais será um tema controverso. Não parece ser possível prever quando esse tipo de opção será atraente. (Resnik 2004, 190)

Como a revisão é proposta contra a "cultura vigente", será raro que isso será feito. Isso parece ser compatível com o modo como a revisão de lógica tem sido feita, em particular destacando que tal revisão não é feita com propósito instrumental da lógica, mas sim em relação ao "cânone da implicação" (Woods 2019).

3. A lógica no pano-de-fundo

O anti-excepcionalismo fornece uma forte alternativa frente à tradicional epistemologia da lógica, mas não é isento de problemas. Em particular, o principal problema para o anti-excepcionalismo ao propor um método de revisão não-excepcional é explicar como esse método pode revisar as leis lógicas sem se basear nessas mesmas leis que estão sujeitas a revisão. Ou, na ausência de tal possibilidade, explicar como isso não é um problema para o anti-excepcionalismo. Esse é o problema da lógica no pano-de-fundo[17]. Tal problema já foi apontado acima em relação a teia-de-crenças de Quine, e ele ressurge também contra a abdução (Woods 2018) e contra o equilíbrio reflexivo (Woods 2019).

O problema da "lógica no pano-de-fundo" se refere a quais princípios lógicos são considerados válidos no contexto de revisão de teorias lógicas. O problema surge quando uma lógica (L) é substituida por outra (L'), mas o ar-

[17] Em inglês, *the problem of the logic in the background*.

gumento usado nessa revisão é um que não é válido em L, "nos encontramos tendo que tomar uma decisão sobre questões sobre a validade lógica para poder prover evidência para uma teoria sobre a validade lógica" (Martin and Hjortland 2020, 15).

Esse problema tem sido usado para sugerir que a lógica seja uma disciplina distinta (excepcional), cuja epistemologia não pode ser compartilhada com as outras ciências, e ameaça seriamente o projeto anti-excepcionalista. Woods (2018, 2019) aponta problemas tanto com o abdutivismo quanto com o equilíbrio reflexivo.

3.1. O problema da oscilação[18]

Na versão do problema do pano-de fundo conta o abdutivismo lógico, Woods (2018) argumenta contra Priest (2016), que já aponta esse problema em relação ao método de revisão que ele próprio sugere. Ao efetuar a revisão de uma lógica em prol de outra, "podemos terminar com uma lógica (aritmética) diferente daquela que nós atualmente usamos. Se fizermos isso, então a computação da escolha deverá ser refeita depois que uma nova teoria é adotada" (Priest 2016, 52).

Priest levanta esse problema mas não o considera seriamente, pois como no cálculo para seleção da melhor teoria há pouca computação feita, esse problema pode nem mesmo aparecer. Dado o problema, ele pondera que é possível que depois de algumas oscilações de escolha alguma estabilidade é alcançada, mas admite que é possível que haja um pior caso em que simplesmente a escolha de teoria oscile sem decisão, já que "aplicar a metodologia abdutiva para avaliar [teorias] em relação a um resultado alternativo resulta em um agente racional oscilando entre a alternativa e a lógica inicial" (Woods 2018, 3). Dado que a escolha de teoria deve ser racional, tanto Priest quanto Woods concordam que tal oscilação deve ser evitada. Ausente de um exemplo não artificial, Priest encerra sua discussão por aí. Tal exemplo é apresentado por Woods (2018), que após expor esse problema, propõe uma solução.

[18] Embora o problema seja apresentado por Woods (2018), essa nomenclatura é adotada somente depois (Woods 2019). Woods (2018) chama esse problema genericamente de "problema da lógica no pano-de-fundo".

O exemplo apresentado por Woods é de uma oscilação na escolha entre a relação de derivabilidade da lógica de núcleo[19] clássica de Neil Tennant, denominada por $\vdash_T$, onde T é não monotônica e não transitiva, e a relação de derivabilidade da lógica clássica denominada por $\vdash_C$. Para utilizar $\vdash_T$na matemática, Tennant demonstra a Eliminação do Corte para T (CET), e como corolário obtém a demonstração da recaptura de C em T (CRT). O problema é que "[n]ão é óbvio que a demonstração de Tennant de CET e CRT evita o uso das propriedades de monotonicidade e transitividade, as quais dever-se-ia mostrar poder ser eliminadas" (Woods 2018, 10). A questão é: considerando as virtudes teóricas de força e informatividade, e dadas as lógicas $\vdash_T$e $\vdash_C$, qual deve ser escolhida como melhor, dada a lógica atual?

Nesse exemplo de uma abdução simples, a força é medida em termos da recaptura do raciocínio matemático, e informatividade é medida em termos de quanta informação é garantida pela demonstração dos teoremas.

> À luz de C, T e C são aproximadamente iguais em relação a força (dado CRT) e T pontua mais do que C em informatividade. No entanto, à luz de T, C domina T em relação a força (já que CRT não está disponível em T), e digamos, T pontua melhor do que C em informatividade. (Woods 2018, 11-12)

Tem-se aí exatamente o caso que Priest precisa para levar esse problema a sério: C sugere a revisão em prol de T, enquanto que T sugere a revisão em prol de C (dado que a força de C é um critério mais relevante do que a informatividade de T). Ocorre um caso simples de oscilação. Dado esse caso, Woods aponta que "a mera possibilidade desses ciclos de decisão teórica já é perturbador o suficiente e aponta um problema profundo com a aplicação da metodologia abdutiva no caso da lógica" (Woods 2018, 19). Ainda assim, ele insiste que:

> [s]e vamos levar a sério a ideia de que devemos escolher uma lógica do mesmo modo que escolhemos teorias científicas, e se nós vemos esse método como um modo de construir justificações para revisão, então precisamos ser capaz de executar a comparação abdutiva sem o sério risco de ciclos de decisões. (Woods 2018, 20)

[19] Em inglês, *classical core logic*.

A solução proposta por Woods é restringir o método de seleção de teorias através do critério do partidarismo[20] lógico, que requer a revisão apenas quando ambas L e L' concordam que uma das teorias é melhor que a outra. A ideia desse critério é permitir que uma teoria seja partidária em relação a ela mesma, de modo a permitir a troca de teoria apenas se a mudança for recomendada por ela própria, e não apenas pela outra teoria proposta.

> A não ser que o resultado de pesar os méritos da minha lógica do pano-de-fundo contra uma alternativa—de um lado à luz da minha própria lógica no pano-de-fundo, e de outro à luz da alternativa proposta—concorde que mudar para a alternativa não é pior do que ficar com minha lógica de pano-de-fundo atual, devemos permanecer com nossa lógica do pano-de-fundo. (Woods 2018, 3)

Esse critério sugere que dada duas lógicas em disputa, a revisão deve ocorrer apenas quando uma das duas teorias é melhor do que a outra, do ponto de vista avaliativo de ambas as opções em disputa. Concedendo que esse procedimento é uma solução plausível para o problema da oscilação (e assim também para o problema da lógica no pano-de-fundo), isso implica em destacar o lugar especial da lógica entre as ciências, já que há um critério extra de seleção para teorias lógicas.

Do ponto de vista abdutivista, a solução do problema da oscilação não deveria vir ao custo de tornar o método de revisão excepcional (isto é, diferente da ciência). No mínimo, seria necessário explicar porque o método de revisão da lógica pode ser excepcional (e que isso não compromete o projeto). Dessa forma, o critério de partidarismo faz apenas o papel parcial em resgatar o projeto abdutivista, mas não faz o suficiente para resguardar esse tipo de anti-excepcionalismo. Alternativas para o abdutivismo são discutidas na seção 4 abaixo.

3.2. O problema dos graus de liberdade

Na versão do problema da lógica no pano-de-fundo contra o equilíbrio reflexivo, Woods (2019) argumenta que esse método de revisão é permissivo demais em relação a como revisar a lógica. Tal liberdade excessiva afeta o

[20] Em inglês, *logical partisanhood*.

equilíbrio reflexivo por este "(a) permitir grande liberdade em como uma teoria será revisada e (b) utilizar uma teoria lógica para avaliar a razoabilidade dos diversos reparos" (Woods 2019, 320). Dada essa liberdade, seria necessário algum tipo de *orientação*[21] em como efetuar a revisão, o que para Woods não pode ser feito satisfatoriamente sem abandonar o anti-excepcionalismo.

Woods explica o equilíbrio reflexivo semelhantemente a Resnik, e misturando o vocabulário utilizado pelos dois, tem-se que tal equilíbrio é atingido quando há *coerência* entre dados e teoria, isto é, tanto a *previsão* de uma teoria está de acordo com os dados quanto os dados são *confirmados* pela teoria. Woods aponta que qualquer definição dos termos utilizado por ele (em itálico acima) pressupõe uma relação de consequência subjacente, que é onde está o problema da lógica no pano-de-fundo. Enquanto que geralmente tais termos são considerados fixos, quando se considera a revisão de lógica, são exatamente eles que estão sujeitos a revisão.

O problema dos graus de liberdade é atribuído a Crispin Wright (1986), em sua crítica a Quine. Woods argumenta que há opções demais em como aumentar a coerência de uma teoria face aos dados. Por exemplo, é possível a princípio: (1) Remover da teoria a parte que está levando a uma previsão indesejável; (2) Remover um dado de consideração; (3) Remover uma previsão, deixando o resto intacto; (4) remover a tensão[22] entre teoria e dado, deixando o resto intacto. Em particular, enquanto que essas últimas duas maneiras de restaurar o equilíbrio funcionam, elas são "grotescas".

> O processo de equilíbrio reflexivo deveria ser um processo no qual correspondemos teoria à evidência e evidência à teoria. Não faz sentido se envolver nesse projeto se nós podemos simplesmente entrar num equilíbrio reflexivo por revisar a teoria excluindo nossas crenças sobre coerência ou predição. (Woods 2019, 330)

Dado esse problema, Woods argumenta que uma solução seria explicar como justificar o ajuste por coerência. Alternativas para isso incluem idealização, invocação de *implicações imediatamente óbvias*, e inclusão de outras virtudes teóricas.

---

[21] Em inglês, *guidance*.

[22] Uma tensão se dá entre uma teoria lógica, uma predição dessa teoria e uma evidência que deveria ser considerada aceita.

Primeiro, a idealização consiste em definir as noções de implicação (no sentido de previsões de uma teoria lógica) e tensão. Caso fosse possível definir idealmente essas noções, não haveria graus demais de liberdade, pois a revisão da teoria estaria já estabelecida de acordo com a definição de tensão e implicação. No entanto, a idealização não está disponível para a revisão da lógica, argumenta Woods, porque a própria possibilidade de idealização requer uma noção de implicação subjacente. Essa solução não está disponível para revisionistas da lógica, pois qualquer tentativa de definir essas noções seria circular.

Segundo, a invocação de implicações imediatamente óbvias consiste em usar uma noção de *profundidade*[23] para orientar as revisões de uma teoria, de modo que a revisão de uma teoria seria orientada pela profundidade de cada pedaço dessa teoria a ser revisada. Não haveria mais graus demais de liberdade, pois a ordem da revisão precisaria apenas seguir a ordem de profundidade. Woods argumenta que tal abordagem é inadequada, pois não há justificativa epistêmica para tratar algumas implicações que são mais óbvias que outras como privilegiadas. Para além disso, a maioria das tensões que ocorrem em teorias lógicas não são óbvias nesse sentido, ou então que são semelhantes em seu grau de "obviedade".

Terceiro, o uso de outras virtudes teóricas leva ao problema da oscilação, da mesma forma que para o uso do método abdutivo. Ainda mais, a suplementação de outras virtudes teóricas (à maneira abducionista) como critérios avaliativos não é suficiente para dar a orientação que Woods propõe. Seria por sua vez necessário alguma orientação ou justificação epistêmica em relação às escolha das virtudes teóricas. Para o abdutivismo lógico, tal justificativa pode ser emprestada das ciências, mas esse movimento não está prontamente disponível para proponentes do equilibrio reflexivo.

Por esses motivos, Woods mantém que as opções para solucionar a versão do problema da lógica no pano-de-fundo que afeta o equilíbrio reflexivo restringem demais o método, talvez até ao ponto de descaracterizar o equilíbrio reflexivo, ou então não são soluções anti-excepcionalistas. Ele conclui

[23] Em inglês, *depth*.

que "[n]ão parece haver uma maneira razoável de salvar o equilíbrio reflexivo como uma metodologia para revisão lógica" (Woods 2019, 338). Woods é assim proponente do método abdutivo, juntamente com o critério de partidarismo já apontado na seção 3.1.

Para Woods, parece ser suficiente que a justificativa para a revisão da lógica seja minimamente distinta do que seria apropriado para as ciências de modo geral para tornar uma proposta já não mais anti-excepcionalista. No entanto, o fato de a "lógica no pano-de-fundo" ser fixa em um dado momento, e assim não sendo apta a revisão nesse momento específico, não a torna imune de revisão. De fato, essa é exatamente a base na qual Resik apoia sua noção de metodologicamente apriori. Talvez essa divergência esteja no âmbito da metodologia científica, ou então esteja no âmbito semântico, no qual a noção de Woods de "anti-excepcionalismo" diverge da apontada aqui. No primeiro caso, não cabe aqui solucionar essa disputa. No segundo, considerando o termo como usado por Woods, a solução que ele propõe na forma de adotar o método abdutivista junto do critério do partidarismo parece divergir do anti-excepcionalismo tanto quanto a aprioridade metodológica de Resnik, e assim talvez o projeto todo anti-excepcionalista não possa sair do chão. Para evitar medidas tão drásticas, parece mais válido aceitar ambas as metodologias do que nenhuma delas.

Para além disso, há ainda dois problemas que podem ser apontados contra o argumento de Woods contra o equilíbrio reflexivo ser um método adequado para o anti-excepcionalismo. O primeiro diz respeito a se existe um problema para início de conversa, e o segundo diz respeito a não haver uma maneira de orientar revisões lógicas usando o equilíbrio reflexivo.

Primeiro, sobre a questão de demais liberdade de revisão, Resnik (2004) já comenta a questão da liberdade de revisão permitida pelo método. Ele não parece se preocupar demais com isso, pois de fato, embora pareça que a lógica é livremente revisável, "[s]erão raros os casos em que alguém irá tentar revisar a lógica" (Resnik, 2004, 190).

Como apresentado na seção 2.2, qualquer proposta de mudança é apresentada à comunidade de lógicos para escrutínio, e pode ser que tais mudanças sejam prontamente aceitas.

> Quando não são, temos uma pequena crise em nossas mãos. Podemos retirar a inferência, ou suspendê-la enquanto tentamos entender porque nossa audiência a rejeitou, ou podemos perseverar na esperança de que eventualmente nosso modo de pensar irá prevalecer; nós podemos ser taxados de loucos e ostracizados. Mas se prevalecermos, pode ser que acabemos por mudar a própria prática. (Resnik 2004, 190)

Quando uma tensão surge, ao mesmo tempo uma proposta de resolução já é apresentada. Então embora a princípio possa haver liberdade demais em como resolver a tensão a nível individual, essa liberdade não se apresenta quando uma revisão é proposta à comunidade, que é o nível em que revisões ocorrem.

Segundo, em criticar a solução da invocação de implicações imediatamente óbvias, Woods parece criticar a noção de metodologicamente apriori que Resnik utiliza, já que definir algumas implicações como mais fundamentais que outras é exatamente o que faz Resnik. No entanto, longe de não haver uma justificativa epistêmica para tal, a "orientação" que Woods indica ser necessária já é oferecida por Resnik. Como já foi apontado na seção 2.2, Resnik já considera algo como o problema da orientação apontado contra Quine, e rejeita a solução quineana, e nisso elabora a noção de metodologicamente apriori. A justificativa epistêmica para manter a lógica no pano-de-fundo fixa é atribuída à comunidade de lógicos, que já operam sob uma determinada metodologia/lógica. Qualquer revisão dessa metodologia, assim como para outras disciplinas normativas, se dá frente a uma comunidade de praticantes.

O fato da justificativa ser normativa, não a torna menos "anti-excepcional" apenas por isso, a não ser que para Woods a "normatividade" deva ser rejeitada no princípio, de modo que a divergência é entre duas noções de lógica diferentes. Isso não parece ser o caso, já que as duas suposições que Woods faz são apenas estas:

> Primeiro, eu suponho que a visões anti-excepcionalistas sobre escolha de teoria lógica são realmente sobre qual lógica adotar como o cânone mais básico da legítima dedução implicativa. [...] Segundo, a metodologia anti-excepcionalista almeja nos prover da habilidade de justificar essa ou aquela revisão lógica. (Woods 2019, 320-321)

Essas duas suposições são completamente compatíveis com a visão de Resnik.

De toda forma, seja o equilíbrio reflexivo um método plausível ou não, está longe de ser um método popular entre anti-excepcionalistas,[24] que consideram essa abordagem apenas marginalmente. Dada a preferência entre anti-excepcionalistas à abdução, o problema da lógica no pano-de-fundo ainda se mantém como um obstáculo a ser superado.

4. Conclusão

O anti-excepcionalismo lógico propõe que teorias lógicas não são justificadas meramente por intuições *a priori*, mas que respondem a evidências. Além disso, propõe um método de revisão para essas teorias que é contínuo com outras disciplinas, sendo nesse sentido anti-excepcional. Duas abordagens distintas podem ser delineadas, uma que lida com a revisão de teorias inteiras, e outra que busca essa revisão passo-a-passo. Como exemplo do primeiro tipo, foi apresentado aqui o abdutivismo lógico, que busca nas ciências naturais o método de revisão para teorias lógicas. Como exemplo do segundo tipo, foi apresentado o método do equilíbrio reflexivo, que aponta o método de revisão da lógica como sendo o apropriado para práticas normativas. Um problema enfrentado por ambos é como propor um método de revisão para a lógica, dado o papel fundacional da lógica. O problema da lógica no pano-de-fundo põe em questão a possibilidade de um método excepcional para a lógica, e é apresentado como o problema da oscilação contra o abdutivismo lógico, e como o problema dos graus de liberdade contra o equilíbrio reflexivo. A solução proposta para o problema da oscilação vem ao custo de propor um critério excepcional para a revisão de teorias, enquanto que para o problema dos graus de

[24] Outra crítica ao equilibrio reflexivo, mais diretamente a visão particular de Resnik, é dada por Shapiro (2000). Embora Shapiro não seja um anti-excepcionalista, pois ao invés de discutir a revisibilidade da lógica ele propõe um tipo de pluralismo lógico, ele também busca articular uma epistemologia alternativa ao que foi chamado aqui de excepcionalismo.

liberdade é dissolvido quando se aceita o papel normativo da lógica, também ao custo de alguma excepcionalidade.

Dadas as duas opções, enquanto que o abdutivismo lógico é um tanto problemático em relação a excepcionalidade do método, o equilíbrio reflexivo é pouco aceito por sustentar que a lógica é normativa.[25] Seria interessante, assim, expandir mais o horizonte do anti-excepcionalismo para outras abordagens. Em particular, parece necessário dar um passo atrás em relação a algumas das suposições feitas pelos abdutivistas.

Um primeiro passo para trás seria rejeitar a posição de que o abdutivismo é o método de revisão de teorias na ciência, haja vista que há amplo respaldo para tal,[26] enquanto se mantém que a lógica e as ciências compartilham um método. Para tal, há amplas opções para serem exploradas na filosofia da ciência. Por exemplo, Diego Tajer (MS) explora como seria um anti-excepcionalismo nos termos dos programas de pesquisa de Thomas Kuhn, e essa seria uma abordagem que entraria na categoria de "comparação de teorias inteiras", nos termos de Woods (2019). Nessa categoria se encaixa também empirismo construtivo de Bas van Fraassen (1980). Ainda relacionando lógica e ciência, pode-se apontar também a possibilidade de utilizar os programas de pesquisa de Imre Lakatos (1978), sendo essa uma abordagem fragmentada sobre a revisão de teorias.

Um segundo passo para trás seria admitir que o método das ciências não é apropriado para a lógica, de modo que a anti-excepcionalidade deve vir de outra forma. O método para a revisão de lógica pode não ser o mesmo das ciências naturais, mas ainda assim pode ser baseado em evidências não *a priori* e haver um método para revisão. A visão de Resnik se encaixaria aqui, pois aponta um método de revisão das disciplinas normativas, e nesse âmbito é possível que outros métodos sejam explorados.

Um terceiro passo (e assim talvez um pulo) seria admitir que a lógica é em certo sentido analítica, isto é, dependente de uma linguagem. Embora essa opção possa parecer tão intragável quanto a visão de Resnik (que admite

[25] Tal normatividade não vem na forma de uma obrigação, mas apenas de regras gerais para uma prática, e não parece que deveria ser antagonizado pelos anti-excepcionalistas mais amplamente.

[26] Por exemplo, ver Van Fraassen (1980, 2001), Douven (2017) e Chakravartty (2017).

um sentido de aprioricidade da lógica), ela é mais aceita na literatura anti-excepcionalista.[27] Nesse sentido, a lógica seria mais como a matemática, e o método adotado seria aquele apropriado a essa disciplina. Nesse sentido, Stephen Read (2018), Priest and Thomason (2007) e G. Russell (2014) relacionam a lógica à teoria de Lakatos da matemática em *Proofs and Refutation* (1976).

Talvez seja necessário desenvolver um método completamente novo, ou então admitir derrota e adotar um "anarquismo" lógico, semelhante a visão de Paul Feyerabend (1993) na ciência. Claro que essa última alternativa não será de todo satisfatória, pois a admissão da falta de método seria o fim da linha para o projeto anti-excepcionalista.

Se para Resnik a lógica é (metodologicamente) apriori, para G. Russell e Read, ela é (metafisicamente) analítica[28] em algum sentido. Os abdutivistas tentam negar categoricamente ambas essas propriedades, e nisso chegam a um certo tipo de naturalismo sobre lógica bastante semelhante a alguma forma de realismo lógico.[29] Tal realismo, por sua vez, parece ser apenas um desvio para chegar ao realismo lógico que já acompanhava a visão excepcional da lógica. Se por um lado o excepcionalismo usa como evidência intuições "claras e distintas" para defender o realismo, o anti-excepcionalismo abdutivistas chega a essa mesma conclusão através de um método naturalista que não abre mão de intuições.[30] De toda forma, a escolha do método apropriado para a revisão da lógica é apenas um reflexo da visão de lógica subjacente, determinado por considerações metafísicas sobre que tipo de fenômeno é a validade lógica.

---

[27] E de fato G. Russell advoga não só abdutivismo lógico, como a analiticidade também. Se tal combinação é plausível é tema para outra discussão.

[28] "Verdadeira em virtude do significado".

[29] Ver Erickson (2020).

[30] Até que ponto essa dissonância metodológica pode ser mantida é um tema para outro estudo.

Agradecimentos

O desenvolvimento deste trabalho foi realizado com apoio da CAPES (processo 88882.376273/2018-01). Agradeço a João Daniel Dantas e a um parecerista anônimo por comentários importantes referentes ao texto.

Referências

Chakravartty, Anjan (2017). "Scientific Realism". Stanford Encyclopedia of Philosophy. Summer 2017 Edition.

Douven, Igor (2017). "Abduction". Stanford Encyclopedia of Philosophy. Summer 2017 Edition.

Erickson, Evelyn (2020). On the metaphysics of (epistemological) logical anti-exceptionalism. *Principia*. No prelo.

Feyerabend, Paul (1993). *Against Method*. London: Verso.

Goodman, Nelson (1955). "The New Riddle of Induction". *Fact, Fiction and Forecast*. Cambridge, Mass.: Harvard UP.

Hjortland, Ole T. (2017a). Anti-exceptionalism About Logic, *Philosophical Studies*, 174(3). http://doi.org/10.1007/s11098-016-0701-8

Hjortland, Ole T. (2017b). Theories of truth and the maxim of minimal mutilation. *Synthese.* http://doi.org/10.1007/s11229-017-1612-8

Hjortland, O. T. (2019). What counts as evidence for a logical theory? *Australasian Journal of Logic* 16(7).

Hjortland, Ole T. and Ben Martin (2020). "Evidence in Logic". Em: M. Lasnon-Aarnio e C. Littlejohn, orgs., *Routledge Handbook of the Philosophy of Evidence*. London, Routledge.

Lakatos, Imre (1976). *Proof and Refutations*: the Logic of Mathematical Discovery. Cambridge: Cambridge UP.

Lakatos, Imre (1978). The Methodology of Scientific Research Programmes. Volume 1 de *Philosophical Papers*. Cambridge: Cambridge UP.

Martin, B. (2020). Identifying logical evidence. *Synthese*. https://doi.org/10.1007/s11229-020-02618-y

Priest, Graham (2006). *Doubt Truth to Be a Liar*. Oxford: Oxford UP.

Priest, Graham e Neil Thomason (2007). 60% Proof: Lakatos, Proof, and Paraconsistency. *Australasian Journal of Logic* 5. http://doi.org/10.26686/ajl.v5i0.1789

Priest, Graham (2016). Logical Disputes and the *a priori*. *Princípios* 23(40).

Quine, W. V. O. (2011). *De um ponto de vista lógico*. Antonio Ianni Segatto, trad. São Paulo: Editora Unesp.

Rawls, John (1971). *A Theory of Justice*. Cambridge, MA.: Harvard UP.

Read, Stephen (2018). Anti-Exceptionalism about Logic. *Australasian Journal of Logic*, 16(7).

Resnik, Michael (2004). "Revising Logic". Em: G. Priest, J.C. Beall, J.C., e B. Armour-Garb, orgs., *The Law of Non-contradiction*. Oxford: Clarendon.

Rodrigues, Cassiano Terra (2011). The Method of Scientific Discovery in Peirce's Philosophy: Deduction, Induction, and Abduction. *Logica Universalis* 5(127). https://doi.org/10.1007/s11787-011-0026-5

Russell, Gillian (2014). Metaphysical Analyticity and the Epistemology of Logic. *Philosophical Studies*, 171(1). https://doi.org/10.1007/s11098-013-0255-y

Shapiro, Stewart (2000). The Status of Logic. Em: P. Boghossian e C. Peacocke, orgs., *New Essays on the a priori*. Oxford: Clarendon. http://doi.org/10.1093/0199241279.003.0014

Tajer, Diego. Kuhnian Anti-exceptionalism about Logic. Manuscrito.

Williamson, Timothy (2007). The Philosophy of Philosophy. Oxford: Blackwell. http://doi.org/10.1002/9780470696675

Williamson, Timothy (2013). Anti-Exceptionalism about Philosophy. *Croatian Journal of Philosophy* 37.

Williamson, Timothy (2017). "Semantic Paradoxes and Abductive Methodology". Em: B. Armour-Garb, org., *Reflections on the Liar*. Oxford: Oxford UP. http://doi.org/10.1093/oso/9780199896042.003.0013

Woods, Jack (2018). Logical Partisanhood. *Philosophical Studies* 176(5). http://doi.org/10.1007/s11098-018-1054-2

Woods, Jack (2019). Against Reflective Equilibrium for Logical Theorizing. *Australasian Journal of Logic* 16(7). https://doi.org/10.26686/ajl.v16i7.5927

Wright, Crispin (1986), "Inventing Logical Necessity". Em: Jeremy Butterfield, org., *Language, Mind, and Logic*. Cambridge, UK: Cambridge UP (1986).

van Fraassen, Bas, (1980). *The Scientific Image*. Oxford: Oxford UP. http://doi.org/10.1093/0198244274.001.0001

van Fraassen, Bas (2001). Constructive Empiricism Now. *Philosophical Studies* 106(1-2). https://doi.org/10.1023/A:1013126824473

O PROBLEMA DOS FUNDAMENTOS DA MATEMÁTICA E AS RAÍZES DO ANTI-EXCEPCIONALISMO LÓGICO

Sanderson Molick[1]

RESUMO

A chamada crise dos fundamentos da matemática, perpetrada pela descoberta dos paradoxos da teoria ingênua de conjuntos, bem como pela descoberta dos teoremas de Gödel, desencadeou uma série de reações à forma como entendemos o conhecimento matemático. Dentre as visões usuais, as mais conhecidas tentam erguer uma base para o conhecimento matemático a partir de elementos a priori, como a noção de prova, ou de "intuição primordial", como é o caso de alguns intuicionistas. O presente artigo tem como objetivo apresentar uma visão antagônica às posições apriorísticas como forma de identificar as raízes históricas do chamado *anti-excepcionalismo lógico*. Para tanto, o texto apresenta as ideias de Lázslo Kalmár e de seu empirismo matemático com o objetivo de contrastar formas de anti-excepcionalismo pré- e pós-Quine. O problema dos fundamentos da matemática é utilizado como critério de demarcação entre estas posições.

Palavras-chave: Anti-Excepcionalismo Lógico. Fundamentos da Matemática. Lázslo Kalmár. Empirismo Matemático.

ABSTRACT

The so-called foundational crisis started by the discovery of the set-theoretic paradoxes in naive set theory, as well as by the discovery of Gödel's incompleteness theorems, was responsible for a series of reactions over the way we understand mathematical knowledge. Among the common views, the most popular ones try to construct mathematical knowledge on a priori elements, such as the notion of proof, or even ideas of "primordial intuitions", as defended by some intuitionists. The purpose of the present paper is to introduce opposite views to the traditional aprioristic ones as a way to trace the historical roots of the so-called *anti-exceptionalism about logic*. In order to do so, the paper presents Lázslo Kalmár's ideas over his mathematical empiricism as a way to contrast two types of anti-exceptionalism, namely pre- and post-

[1] Doutorando em Filosofia pela Universidade Federal do Rio Grande do Norte/UFRN e a Ruhr-Universität Bochum/RUB.
E-mail: smolicks@gmail.com. ORCID: 0000-0003-2918-1055.

Quine. The problem of mathematical foundations is explored as a demarcation criterion between these positions.

Key-words: Logical Anti-Exceptionalism. Foundations of Mathematics. Lázslo Kalmár. Mathematical Empiricism.

1. Introdução

O *anti-excepcionalismo lógico* afirma que a lógica não é uma exceção entre as ciências empíricas. Os métodos de teste e decisão das teorias e proposições lógicas são semelhantes aos métodos de teste e decisão das teorias científicas. Com isso, de acordo com o anti-excepcionalismo, a evidência empírica também desempenha um papel importante na justificação dos princípios lógicos.

O presente artigo examina as raízes históricas do anti-excepcionalismo lógico. Embora o termo anti-excepcionalismo tenha sido cunhado apenas recentemente por Timothy Williamson [Williamson, 2013], a ideia de que a lógica não difere das ciências empíricas remete também aos empiristas matemáticos da primeira metade do século XX. Dentre eles, encontramos autores como John Stuart Mill, Bertrand Russell, Fraenkel, Georg Kreisel e Lázslo Kalmár.

O artigo divide o anti-excepcionalismo historicamente em duas fases. Como critério de demarcação entre os anti-excepcionalistas da primeira fase e os anti-excepcionalistas atuais, o artigo utiliza o problema dos fundamentos e da justificação do conhecimento matemático como forma de sublinhar as diferenças entre as duas formas de anti-excepcionalismo. Portanto, o uso do termo anti-excepcionalismo estará ligado ao uso e aceitação de evidência a posteriori como forma de justificação do conhecimento, seja do nosso conhecimento dos princípios da lógica, ou do conhecimento das verdades matemáticas.[2] No percurso de desenvolvimento do artigo, a expressão *problema dos fundamentos* se referirá ao problema de encontrar fundamentos apriorísticos para o conhecimento matemático. Programa iniciado a partir dos esforço logicista de Frege e

[2] Note, contudo, que o artigo não pretendo colapsar o empirismo matemático ao anti-excepcionalismo lógico, mas apenas contrastar ambas posições a partir de um ponto em comum, a saber, da aceitação da evidência empírica como forma de justificação de suas verdades.

sucedido por diversos programas de pesquisa, como o formalismo de Hilbert e o intuicionismo de Brouwer.

Na primeira forma de anti-excepcionalismo, aqui denominada de *anti-excepcionalismo pré-quineano*, o caráter não-apriorístico da lógica implica repensar o problema dos fundamentos da matemática a partir da admissão de que certas áreas da matemática possuem validação empírica. Na segunda fase, aqui denominada de *anti-excepcionalismo pós-quineano*, o caráter não-apriorístico da lógica é utilizado como via de solução para o problema da rivalidade entre lógicas, de modo a fazer uso de modelos abdutivos como processo de decisão entre teorias lógicas rivais.

Uma forma de anti-excepcionalismo pré-quineano é ilustrada a partir de Lázslo Kalmár, autor central na defesa e formulação do empirismo matemático como via de solução ao problema dos fundamentos. Para tanto, o artigo apresenta uma descrição da filosofia da matemática de Kalmár e o situa em comparação às formas contemporâneas de anti-excepcionalismo lógico.

O artigo possui a seguinte estrutura: a seção 2 introduz o problema dos fundamentos e discute os problemas filosóficos que nascem desta; na seção 3 é introduzido o anti-excepcionalismo lógico. As diferentes formas de anti-excepcionalismo são classificadas em anti-excepcionalismos pré- e pós-quineano. A distinção entre ambas formas de anti-excepcionalismo é introduzida a partir do modo como ambos vêem o problema dos fundamentos. Por fim, na seção 4, o anti-excepcionalismo de Kalmár é apresentado como exemplo paradigmático de anti-excepcionalismo pré-quineano. Os aspectos centrais de sua filosofia da matemática são apresentados e discutidos.

2. O problema dos fundamentos

A lógica contemporânea se desenvolve do esforço logicista de fundamentar o conhecimento da aritmética nas leis e regras da lógica. É a partir dos trabalhos de Frege, em especial, em *Os fundamentos da aritmética,* que encontramos uma formulação dos axiomas e regras da, hoje chamada, lógica clássica. Se encontramos em Kant a defesa de que o conhecimento matemático se origina nas intuições puras do tempo e do espaço, há em Frege a tentati-

va de se demonstrar que o conhecimento da aritmética pode ser fundamentado na lógica. O programa logicista, conforme já conhecido, culmina na descoberta do paradoxo de Russell e desencadeia a chamada *crise dos fundamentos*.

Duas questões se tornam centrais para a filosofia da lógica e da matemática a partir da crise dos fundamentos, a saber: 1) qual a fonte do conhecimento matemático?, e 2) como podemos justificar o conhecimento matemático ou, em particular, como podemos justificar nosso conhecimento da aritmética? Uma das vias de solução buscou responder a segunda questão a partir de métodos apriorísticos, isto é, utilizando os métodos da lógica para demonstrar como nosso conhecimento da matemática pode ser reduzido a princípios lógicos fundamentais. Para eles, mostrar que o conhecimento matemático pode ser fundado em princípios lógicos significava demonstrar o caráter infalível da matemática. Assim, a matemática fixaria seu lugar como uma disciplina dotada de um grau de certeza maior que as ciências empíricas, isto é, a matemática seria uma *exceção* dentre as outras ciências.

Uma outra via de solução, aqui chamada de *anti-excepcionalista*, enxergava o conhecimento matemático enquanto revisável e passível de refutação empírica. Para muitos destes autores, os resultados matemáticos jamais poderiam ser assentados sobre princípios sólidos e indubitáveis. Em [Lakatós, 1976], encontramos um registro de alguns dos autores que partilhavam esta visão, como Fraenkel, Church, Weyl, dentre outros. Tal visão sobre o conhecimento da matemática era denominada de *empirismo matemático*. No que segue, apresento o empirismo matemático tal como encontrado na visão de um autor central deste período, Lázslo Kalmár.

3. A solução anti-excepcionalista

Em sua acepção moderna, o anti-excepcionalismo lógico é caracterizado por [Hjortland, 2016] a partir de duas teses centrais: 1) a lógica não é a priori; e 2) as teorias lógicas podem ser escolhidas a partir do mesmo processo de decisão utilizado para escolher entre teorias científicas rivais, isto é, a partir da comparação das virtudes epistêmicas de cada teoria. Dentre estas

virtudes, encontramos a consistência, a adequação aos dados, o poder explicativo, a fecundidade etc.

No presente artigo dividirei os anti-excepcionalistas a partir de dois momentos centrais: os anti-excepcionalistas da *era pré-quineana* e os anti-excepcionalistas da *era pós-quineana*. Dentre os do segundo grupo, destaco [Williamson, 2013], [Priest, 2016] e [Hjortland, 2016]. Já entre os primeiros, é possível classificar autores que vão desde John Stuart Mill à Georg Kreisel, Lazslo Kalmár e Imre Lakatos. Contudo, por razões de espaço (e do fato de que ainda é difícil o acesso à obra de alguns destes outros autores), focarei apenas em Kalmár.

Influenciados pelos trabalhos de Quine e pelas questões da filosofia da lógica contemporânea, tais como a rivalidade entre lógicas, o pluralismo lógico e a validade dos princípios da lógica clássica, os trabalhos dos anti-excepcionalistas da era pós-quineana possuem maior foco na tese (2) acima e em como os métodos de escolha entre teorias científicas rivais podem ser aplicados na lógica. A escolha de Quine como ponto de inflexão está no fato de que tal autor se enquadra como um dos primeiros a utilizar critérios abdutivos como forma de defender a superioridade da lógica clássica em relação à algumas lógicas não-clássicas.

Em contrapartida, os anti-excepcionalistas da era pré-quineana se caracterizam por uma maior preocupação na tese (1), isto é, em como caracterizar adequadamente o caráter não-apriorístico da lógica e no que isto representaria para a forma como justificamos o conhecimento matemático. Deste modo, enquanto os anti-excepcionalistas da era pós-quineana acreditam que encontrar um fundamento adequado para a matemática significa encontrar a melhor teoria lógica dentre as concorrentes, os anti-excepcionalistas da era pré-quineana acreditam que o caráter não-apriorístico da lógica, por si só, demonstra que a matemática não precisa de uma teoria lógica para a justificação de seus princípios.

Os anti-excepcionalistas da era pós-quineana encontram no caráter não-apriorístico da lógica uma via de saída para o problema da rivalidade entre lógicas. Estes autores ainda reconhecem a necessidade de encontrar funda-

mentos adequados para as teorias matemáticas, para as teorias da verdade, as teorias do significado, etc. Assim, se encontramos uma variedade de sistemas lógicos disponíveis e nos perguntamos qual destes deveria fundamentar certas áreas do conhecimento, os anti-excepcionalistas da era pós-quine defendem que podemos escolher cada sistema a partir do uso de modelos abdutivos de escolha.

É importante ressaltar, contudo, que a noção de evidência utilizada pelos autores da era pré-quineana não é a mesma da era pós-quineana. Embora seja verdade que ambos os grupos aceitam o papel da evidência a posteriori na lógica e na matemática, é somente nos autores da era pós-quineana que encontramos a aceitação da evidência psicológica e de outras formas de evidência baseadas em dados não-observacionais. Uma das razões para tal diferença entre as concepções de evidência reside no fato de que as discussões contemporâneas acerca da falibilidade de certos princípios da lógica clássica parecem aceitar uma ampla gama de dados como forma de justificar determinadas lógicas não-clássicas. Desse modo, é comum encontrarmos dados diferentes como situações epistêmicas, construções matemáticas, discussões metafísicas etc, como forma de legitimar a falibilidade de um princípio clássico em favor de sua revisão[3].

Conforme mecionado, encontramos em Kalmár, e em certa medida também em Lakatos, a crença de que o caráter não-apriorístico da lógica revela apenas que o problema de encontrar fundamentos seguros para a matemática não se coloca da maneira usual, pois os axiomas e os princípios lógicos partem da experiência empírica e são validados por esta. É verdade que, por razões históricas óbvias, estes autores não lidam com o problema da rivalidade entre lógicas. Contudo, é importante ressaltar que, diferentemente dos autores da era pós-quineana, a crença no caráter empírico da lógica conduz estes autores à uma visão distinta sobre o problema dos fundamentos. Na seção seguinte, será exposta a visão de Kalmár acerca do conhecimento matemático.

[3] Para mais sobre este ponto, ver [Molick, 2020].

4. O anti-excepcionalismo de Kalmár

Conforme discutido, as raízes do anti-excepcionalismo lógico se encontram no *empirismo matemático* defendido por muitos filósofos e matemáticos da primeira metade do século XX. De acordo com o empirismo matemático, as verdades matemáticas são justificadas através da evidência empírica, assim como nas outras ciências. Uma figura que desempenha papel central na defesa e formulação de muitas das teses do empirismo matemático é o lógico húngaro Lazslo Kalmár (1905-1976).

Em [Kalmar, 1967], o autor argumenta que a invenção da abstração, da inferência dedutiva, e do método axiomático, foi responsável por grande parte do desenvolvimento da matemática. Contudo, tais métodos levaram os matemáticos a entender a matemática como uma disciplina "puramente dedutiva" em oposição às disciplinas que fazem uso de testes empíricos. De acordo com Kalmár, o método da *abstração* possui as seguintes funções centrais: 1) descobrir regularidades gerais entre os fatos empíricos, e 2) permitir a construção de conceitos idealizados, tais como os de ponto ou reta, para os quais existem apenas aproximações no mundo real, mas cuja simplicidade lógica os torna mais adequados para o uso e manipulação na inferência dedutiva.

Para Kalmár, a abstração é o primeiro passo em direção aos métodos de produção dos teoremas e resultados matemáticos. É a partir da abstração que conseguimos a via de acesso à axiomatização, conceitualização e até mesmo à construção de modelos. O percurso do conhecimento matemático é descrito por Kalmár em [Kalmár, 1942]. Conforme descreve o autor, o conteúdo de nossas intuições permite o matemático alcançar passos de abstração cada vez maiores. Assim, o conhecimento matemático parte da realidade empírica e produz intuições que serão refinadas a partir da conceitualização e axiomatização. É somente a partir daí que o conteúdo dos axiomas será abstraído de forma a produzir teorias matemáticas distantes da intuição. Tais teorias serão reguladas por princípios e regras básicas e é a partir delas que o matemático construirá diferentes modelos de teorias formais. Dentro deste percurso, a busca por rigor se distancia e sacrifica cada vez mais o papel das intuições que antes originaram a teoria. É este o ponto central da crítica de Kalmár – o fato

de que, ainda que as teorias matemáticas possam se reduzir a jogos formais, elas se baseiam em intuições que estão calcadas no mundo empírico. Deste modo, entender a matemática como uma ciência puramente dedutiva levou os matemáticos a esquecer que os axiomas e regras foram abstraídos de regularidades empíricas, as quais já possuem validade porque assim foram confirmadas em nossas práticas inferenciais.

Em [Kalmár, 1962], Kalmár argumenta que vários fatores contribuíram para sustentar a visão da matemática enquanto ciência pura dedutiva, a saber: 1) a consolidação de uma teoria do a priori, segundo a qual não é necessário o teste de axiomas e regras de inferência intuitivamente evidentes; 2) a visão popular de que a matemática é uma disciplina "infalível" (o que neste contexto significa *não-revisável*); 3) a ideia de que a matematica serve como modelo de ciência a ser buscado pelas outras áreas do conhecimento em razão de sua precisão; 4) o próprio sucesso da matemática e a extensão da aplicação do método de abstração, os quais levaram a descoberta de resultados e teoremas matemáticos cada vez mais abstratos.

Deste modo, como é possível confirmar, encontramos em Kalmár a ideia de que o método de justificação do conhecimento matemático não se reduz aos métodos a priori. Embora a matemática faça uso de procedimentos a priori na demonstração de seus resultados, o recurso primário do conhecimento matemático é a experiência empírica, a qual será abstraída para a descoberta de regularidades e generalizações. Conforme ressalta o autor: "(…) *Por que não confessamos que a matemática, assim como as outras ciências, baseia-se, e deve ser testada, na prática [experiência]*?"[4]

Neste ponto, embora Kalmár aponte para o uso da evidência a posteriori no papel de justificação do conhecimento matemático, não é claro que tipo de evidência é aceita por Kalmár ou como ocorreria o processo de revisão das verdades matemáticas a partir de evidências empíricas[5]. Mais ainda, embora o autor forneça uma breve descrição do processo de evolução da empiria à abstração [Kalmár, 2011], assim fornecendo informações sobre sua epistemolo-

[4] [Kalmár, 1962, p. 7], tradução minha.

[5] O autor reconhece, contudo, a necessidade de abordar tais questões e as menciona como problemas em aberto para um programa de pesquisa que busque desenvolver fundamentos empíricos para os fundamentos da matemática.

gia da lógica, o mesmo não parece fornecer uma descrição dos mecanismos heurísticos do processo de revisão de teorias matemáticas[6].

É com base em sua visão acerca da justificação do conhecimento matemático que Kalmár formulará sua crítica à obsessão formalista de encontrar fundamentos puramente dedutivos para a matemática. Para o autor, a consistência de muitas de nossas teorias formais é apenas "um fato empírico" garantido pelo fato de que se fundamentam em axiomas abstraídos do caráter empírico de nossas práticas inferenciais. Com isso, a busca por fundamentos puramente dedutivos acaba por esquecer o lugar de onde tais axiomas e regras de inferência foram abstraídos, isto é, dos fenômenos empíricos. Assim, a busca pelos fundamentos da matemática só faz sentido na medida em que se reconheça que muitos princípios matemáticos são validados empiricamente e podem ser revisados à luz de nova evidência. Para Kalmár, mesmo proposições como a tese de Church deve ser investigada como um problema empírico, o qual é passível de teste e revisão (ver [Kalmár, 1959]).

4.1. Os fundamentos da matemática – para onde?

A crítica de Kalmár em [Kalmár, 1962] resume-se em seu slogan principal: *fundamentos da matemática – para onde a partir de agora?*[7] Para ele, o programa de pesquisa em torno dos fundamentos da matemática parte de premissas errôneas. Apesar da existência de diversos programas de pesquisa envolvidos na fundamentação da matemática, como por exemplo o intuicionismo, a teoria de tipos de Whitehead e Russell, e a teoria da prova hilbertiana, ainda se faz necessário desenvolver novas linhas de pesquisa a partir da admissão de que a matemática não é uma ciência pura dedutiva.

A saída apontada pelo autor é reconhecer a matemática enquanto ciência com fundamentos empíricos. Uma das consequência de tal admissão faria com que fosse possível admitir o raciocínio indutivo dentro da prática matemática. Para Kalmár, o raciocínio indutivo e o raciocínio dedutivo possuem alguns problemas intrínsecos. Por exemplo, ainda que o raciocínio induti-

[6] É possível encontrar a descrição de tais mecanismos heurísticos, bem como do tipo de evidência aceita, em [Lakatos, 1976].

[7] No original "*Foundations of mathematics – whither now?*"

vo sofra dos problemas usuais ligados ao fato de que não podemos garantir a verdade em todos os casos de uma afirmação universal, as afirmações existenciais demonstradas por dedução também sofrem dos problemas comuns apontados pelo intuicionistas.

Uma vez admitidos os fundamentos empíricos da matemática, ainda se seguiriam novos problemas ligados aos mecanismos heurísticos de confirmação e refutação dos teoremas matemáticos. Dentre eles, estaria o problema de definir quais são os "observáveis" aos quais as proposições matemáticas estariam sujeitas à confirmação ou refutação, bem como o problema de delimitar quais proposições matemáticas estariam sujeitas à testes empíricos e quais não estariam[8]. É a partir de tais questões que Kalmár avança parte dos problemas que farão parte da agenda atual dos anti-excepcionalistas lógicos. Aceitar o papel da evidência a posteriori dentro do processos de justificação das teorias lógicas e matemáticas leva aos problemas centrais enfrentados pelos anti-excepcionalistas da era pós-quineana.

5. Considerações finais

No presente artigo examinei o anti-excepcionalismo lógico a partir do problema dos fundamentos da matemática. Com isto, foi defendida uma classificação das formas de anti-excepcionalismo pré- e pós-quineana como meio de ressaltar as divergências entre os anti-excepcionalistas da segunda metade do século XX e os anti-excepcionalistas da filosofia da lógica atual. O artigo focou unicamente no pensamento de Lázslo Kalmár. Contudo, as comparações feitas aqui poderiam ser exploradas entre diferentes autores do empirismo lógico. Assim, permanece como trabalho futuro o desenvolvimento de uma investigação histórica de outros autores que compuseram os anti-excepcionalistas da fase pré-quineana. Também se faz necessário uma comparação do trabalho destes autores em torno de problemas como o de definir a noção de evidência empírica para a lógica e a matemática, a heurística de confirmação e refutação de tais teorias à luz de evidências a posteriori, bem como o

[8] O problema de identificar os observáveis de confirmação e/ou refutação das proposições matemáticas será posteriormente tratado por Lakatos em *A renaissance of empiricism in the recent philosophy of mathematics* (The British Journal for the Philosophy of Science, 1976).

papel de tais elementos em debates como o pluralismo lógico e o processo de escolha de teorias lógicas.

6. Referências

[Hjortland, 2016] Hjortland, O. T. (2016). Anti-exceptionalism about logic. Philosophical Studies, pages 1–28.

[Kalmár, 1959] Kalmár, L. (1959). An argument against the plausibility of Church's Thesis. In: Constructivity in Mathematics, North Holland-Amsterdam.

[Kalmar, 1967] Kalmar, L. (1967). Foundations of mathematics—whither now? In Studies in Logic and the Foundations of Mathematics, volume 47, pages 187–207. Elsevier.

[Kalmár, 2011] Kalmár, L. (2011). The development of mathematical rigor from intuition to axiomatic method. In Der Wiener Kreis in Ungarn/The Vienna Circle in Hungary, pages 269–288. Springer.

[Lakatos, 1976] Lakatos, I. (1976). A renaissance of empiricism in the recent philosophy of mathematics. The British Journal for the Philosophy of Science, 27(3):201–223.

[Molick, 2020] Molick, S. (2020). Logical theory revision through data underdetermination: an anti-exceptionalist exercise. A publicar em Principia: an international journal of epistemology.

[Priest, 2016] Priest, G. (2016). Logical disputes and the a priori. Princípios: Revista de Filosofia, 23(40):29–57.

[Szabó, 2018] Szabó, M. (2018). Kalmár's argument against the plausibility of Church's thesis. History and Philosophy of Logic, 39(2):140–157.

[Williamson, 2013] Williamson, T. (2013). Modal logic as metaphysics. Oxford University Press.

TOLERANCIA A LA INCONSISTENCIA EN LAS CIENCIAS SOCIALES: CONTRADICCIONES ENTRE TEORÍA Y OBSERVACIÓN EN LA ECONOMÍA

María del Rosario Martínez-Ordaz[1]

Josafat Iván Hernández-Cervantes[2]

RESUMEN

Aquí defendemos que el estudio de la tolerancia a la inconsistencia en las ciencias sociales –específicamente, la tolerancia a las contradicciones entre teoría y observación–, puede ser revelador con respecto a los mecanismos que subyacen al uso racional de la información inconsistente. Para hacerlo, ofrecemos una tipología de contradicciones de este tipo. Enfatizamos algunos de los aspectos destacados de estas contradicciones, y defendemos que dichos aspectos tienen un papel importante para la selección de los mecanismos inferenciales que permiten la tolerancia a las contradicciones. Ilustramos esto con casos de economía.

Palabras-clave: Tolerancia a la Inconsistencia. Contradicciones Entre Teoría y Observación. Economía.

RESUMO

Defendemos aqui que o estudo da tolerância à inconsistência nas ciências sociais - especificamente, a tolerância às contradições entre teoria e observação - pode ser revelador no que diz respeito aos mecanismos que fundamentam o uso racional de informações inconsistentes. Para fazer isso, oferecemos uma tipologia de contradições desse tipo. Enfatizamos alguns dos aspectos salientes dessas contradições, e defendemos que esses aspectos desempenham um papel importante na seleção de mecanismos inferenciais que permitem a tolerância às contradições. Ilustramos isso com casos econômicos.

Palavras-chave: Tolerância à Inconsistência. Contradições Entre Teoria e Observação. Economia.

[1] Programa de Pós-Graduação em Filosofia, Universidade Federal do Rio de Janeiro. E-mail: martinezordazm@gmail.com. ORCID: 0000-0003-2118-3515.

[2] Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México. E-mail: josafativan.hernandez@gmail.com. ORCID: 0000-0002-8077-5900.

ABSTRACT

Here, we argue that the study of inconsistency toleration in the social sciences -particularly the toleration of contradictions between theory and observation-, can be revealing about the mechanisms that underlie the rational use of inconsistent information. In order to do so, we provide a typology of contradictions between theory and observation. We emphasize some of the most important features of these contradictions and we argue that they play a significant role for determining the inference mechanisms that allow for the rational toleration of contradictions. We illustrate this with cases from economics.

KEYWORDS: Inconsistency Toleration. Contradictions Between Theory and Observation. Economics.

1. Introducción

Durante las últimas décadas, se ha asumido que el estudio de las creencias contradictorias, y de las distintas maneras en las que resolvemos las contradicciones, juega un papel fundamental para tanto la elucidación de las bases de la racionalidad humana (cf. ROVANE, 2004), como para la mejor comprensión del estatus de algunos de los principios lógicos mejor aceptados.

Por un lado, una *contradicción* es un par de proposiciones del cual una es la negación de la otra, y puede ser representada de la forma: α y $\neg\alpha$. El *Principio de Explosión* (PE) dice que si cualquier teoría cerrada bajo una relación de consecuencia lógica explosiva contiene una contradicción, entonces esta teoría es trivial. Una teoría (formal) es *trivial* si cualquier proposición es un teorema de la misma. Por tanto, cualquier teoría que contenga una contradicción y esté cerrada bajo una relación de consecuencia lógica explosiva, será trivial. Esto ha causado la impresión de que la relación que existe entre las contradicciones y la trivialidad es tan estrecha que generalmente cuando se encuentra una contradicción en una teoría, la trivialidad de la misma está garantizada. Por otro lado, desde un punto de vista epistemológico, la adopción de creencias mutuamente contradictorias suele ser indicativa de alguna suerte de irracionalidad (Cf. MOORE 1942: 543). Aquellos agentes que adoptan una creencia α y luego, sin haberla abandonado previamente, adoptan su negación enfrentan, al menos, dos retos: (1) dar cuenta de la justificación y su mutua in-

dependencia que se tiene para ambas creencias, y (2) evitar la trivialización de su razonamiento al realizar inferencias a partir de (o con) la información inconsistente en cuestión. A la luz de lo anterior, la detección y el análisis de las contradicciones han tenido una larga historia dentro del pensamiento filosófico occidental, ya sea que la motivación de fondo esté inspirada en la irracionalidad asociada a poseer redes inconsistentes de creencias o en el supuesto carácter trivializador de las contradicciones. Y he ahí la importancia de estudiar el fenómeno de la tolerancia a las contradicciones desde la intersección entre la lógica y la epistemología.

Tanto en lógica como en epistemología, han existido diversos programas de investigación dedicados al estudio de las formas en las cuales es posible trabajar con información inconsistente sin poner en peligro la racionalidad de los agentes involucrados. Asumiendo que las diversas prácticas epistémicas en las ciencias satisfacen altos estándares de racionalidad, y que por tanto, su estudio resultará revelador con respecto a las mejores estrategias para la preservación de la racionalidad en contextos problemáticos; lógicos y epistemólogos han centrado su atención en aquellas prácticas que resultan ser más útiles para el manejo de información inconsistente en las ciencias (cf. BATENS 2002, 2017; BUENO 2006, 2017; CARNIELLI y CONIGLIO 2016: Chap. 9; ESTRADA-GONZÁLEZ 2014; MEHEUS 2002). Sin embargo, la mayoría de estos estudios se han centrado en ciencias fuertemente formales, como la matemática y la física, en especial, en las contradicciones internas a las teorías científicas más importantes de estas disciplinas (Cf. PRIEST 2002, FRISCH 2004, BUENO 2006, MARTÍNEZ-ORDAZ 2017).

En lo que sigue, buscamos extender el dominio de interés tanto a otros tipos de contradicciones como a otras disciplinas científicas. En particular, defendemos que el análisis de las prácticas de tolerancia a las contradicciones entre teoría y observación en las ciencias sociales puede resultar revelador sobre el uso racional de información inconsistente. Ofrecemos una tipología de contradicciones de este tipo, a través de la cual enfatizamos aspectos importantes de las mismas que pueden tener un efecto importante en los mecanismos inferenciales asociados a su tolerancia.

El artículo está dividido en seis secciones. En la Sec. 2, presentamos el debate sobre la plausibilidad de que la ciencia sea inconsistente; para hacerlo, introducimos las dos posiciones filosóficas en contienda, la visión *tradicional* y la *visión tolerante*. Aquí mismo, presentamos uno de los argumentos mejor conocidos de la visión tolerante y discutimos las debilidades de dicho argumento. En la Sec. 3, explicamos la relación que existe entre la tolerancia a las contradicciones en las ciencias y la ignorancia de valores de verdad, así como la relación entre casos de tolerancia a las contradicciones y herramientas paraconsistentes. En la Sec. 4, argumentamos a favor del estudio de las contradicciones entre teoría y observación para promover una mejor comprensión de la tolerancia a la inconsistencia en las ciencias. En la Sec. 5, presentamos una tipología de las contradicciones entre teoría y observación que es posible encontrar en dichas disciplinas científicas y ofrecemos dos ejemplos de la economía que muestran cómo los economistas pueden trabajar con contradicciones para estudiar fenómenos complejos. Finalmente, en la Sec. 6, ofrecemos algunas reflexiones sobre la importancia de incorporar las ciencias sociales al dominio de estudio del razonamiento inconsistente no-trivial.

2. Sobre la plausibilidad de que la ciencia sea inconsistente

Esta sección está dedicada a abordar los debates con respecto a la plausibilidad de la tolerancia a la inconsistencia en las ciencias. Para hacerlo, la sección está dividida en dos partes: en Sec. 2.1, presentamos algunos preliminares que son fundamentales para comprender dicho debate. En Sec. 2.2 ofrecemos una descripción general de las dos posiciones más robustas en el debate, la *visión tradicional* y la *visión tolerante,* y discutimos una de las objeciones más importantes a la metodología de la visión tolerante.

2.1. Preliminares

Una relación de consecuencia lógica es *paraconsistente* si no es explosiva, es decir, si no valida el Principio de Explosión (PE). Y una teoría formal es paraconsistente si, a pesar de contener una contradicción, no es trivial.

Tolerancia a la inconsistencia es la práctica de poder razonar ya sea a partir de o con información inconsistente, y seguir siendo capaz de identificar las consecuencias de nuestro razonamiento que son sensatas en determinados contextos de las que no lo son. Por esta razón, generalmente, la tolerancia a la inconsistencia se ha visto como un indicador (confiable) de la evasión de la explosión lógica, a pesar de tener información inconsistente en juego.

La tolerancia a la inconsistencia en las ciencias puede ocurrir al menos en tres niveles distintos:

- **A nivel de los compromisos doxásticos de los agentes epistémicos:** Este nivel de tolerancia puede ocurrir cuando los agentes epistémicos creen que ambas, α y $\neg\alpha$, son verdaderas pero tienen tanto explicaciones racionales independientes para cada una, como estrategias de compartimentalización que impiden que el agente las use "juntas". Generalmente, el análisis de los compromisos doxásticos de los científicos ha sido abordado por epistemólogos e historiadores de la ciencia, arrojando como resultado que existe una fuerte ausencia de evidencia a favor de este nivel de tolerancia en las ciencias (Cf. VICKERS Cap 3-8).[3]
- **A nivel del razonamiento de agentes epistémicos específicos:** Un agente es tolerante a las contradicciones si, independientemente de sus compromisos doxásticos logra razonar a partir de o con información inconsistente sin poner en riesgo su racionalidad. Este nivel de tolerancia ha sido explorado por diversos lógicos no-clásicos que han ofrecido, al menos, dos explicaciones alternativas a cómo es esto posible: por un lado, se ha sugerido que existen fragmentos del razonamiento científico que no están cerrados bajo una relación de consecuencia lógica que sea explosiva, y que por tanto, permiten la presencia de cier-

[3] En la literatura se ha sugerido que los únicos casos que podrían exhibir tolerancia a las contradicciones en este nivel, son aquellos donde los agentes están completamente comprometidos con la verdad de ambas proposiciones. Es decir, casos donde el agente cree que α es verdadera y que $\neg\alpha$ es una idealización (i. e., es falsa) que solamente resulta útil en algún contexto, no son considerados instancias legítimas de tolerancia a las contradicciones (Cf. VICKERS 2013: Cap. 2, DAVEY 2014).

tas contradicciones sin trivializar de dicho fragmento de razonamiento (Cf. MEHEUS 2002). Por otro lado, se ha defendido que los agentes humanos disponen de estrategias de razonamiento que les permiten separar (o compartimentalizar) cada uno de los componentes de la contradicción, de tal forma que evitan que haya comunicación entre ellos y con ello, previenen la explosión lógica (Cf. BROWN y PRIEST 2004, SWEENEY 2013).

- **A nivel de la teoría:** Una teoría es tolerante a las contradicciones si, a pesar de contener al menos una contradicción, es no trivial -independientemente de que exista un agente que razone a partir de dicha teoría. En algunos casos, las teorías son tolerantes a las contradicciones porque están cerradas bajo una relación de consecuencia lógica paraconsistente; ejemplos de esto son algunas de las reconstrucciones de la teoría de conjuntos fregeana (e.g., BATENS 2019).

En lo que sigue, y hasta indicar lo contrario, nos concentramos en las contradicciones a nivel de las teorías científicas –y las prácticas inferenciales asociadas a su uso.

Asumamos que una teoría científica *T* puede ser caracterizada de la siguiente manera: $T =< D, R_n^i >$, donde *D* indica un dominio de aplicación específico y R_i es una familia de relaciones n-arias que valen entre elementos de *D* (cf. BUENO 1997: 588). *T* es un conjunto de substructuras parciales $< A', A'', \ldots A^n >$ de la forma $A =< D, R_k >_{k \in K}$. R_k indica relaciones parciales, disjuntas, tales que $R_k =< R_{k1}, R_{k2}, R_{k3} >$, donde R_{k1} es el conjunto de n-tuplas que se asume pertenecen a R, R_{k2} es el conjunto de n-tuplas que se cree no pertenecen a R y R_{k3} es el conjunto de n-tuplas que aún no ha sido determinado si pertenecen o no a R (ver BUENO y FRENCH 2011: 858–859).[4]

[4] Estamos conscientes de que en la actualidad existe la latente discusión sobre qué forma de representar nuestras teorías científicas es más adecuada y, epistémicamente, más útil para describir las prácticas epistémicas de los científicos, sin embargo, considerando los fines de este artículo creemos que esta caracterización será suficiente. Si el lector está interesado en un análisis exhaustivo sobre las distintas caracterizaciones de teoría empírica, ver: [HALVORSON 2016].

Una teoría *T* será inconsistente si dadas dos subestructuras parciales A' y A'', una contiene la proposición *s* y la otra contiene su negación –y ambas proposiciones se asumen como, al menos, parcialmente verdaderas.

2.2. La visión tradicional y la visión tolerante

¿Pueden ser nuestras teorías científicas inconsistentes y al mismo tiempo no poner en riesgo la racionalidad científica? Esta es la pregunta central del debate sobre la plausibilidad de la tolerancia a la inconsistencia en las ciencias. Con la intención de responder a tal interrogante, dos posiciones, han surgido dos posiciones distintas que podemos llamar: *la visión tradicional* y *la visión tolerante*.

Por un lado, de acuerdo con *la visión tradicional*, las contradicciones en las ciencias, tanto en las teorías como en las prácticas científicas, han sido caracterizadas como problemáticas por al menos tres razones. Primero, de estar las teorías científicas cerradas bajo una relación de consecuencia lógica explosiva, si contuvieran una contradicción, dejarían inmediatamente de ser confiables, dado que cualquier cosa y su negación podrían ser ambas predicciones legítimas de dichas teorías (cf. HEMPEL 2000: 79; POPPER 1959). Segundo, de estar la racionalidad humana cerrada bajo una relación de consecuencia lógica explosiva, la racionalidad de los científicos estaría en grave peligro si sus teorías o modelos predilectos fueran inconsistentes; en particular, todas sus creencias en favor de sus teorías serían irracionales (cf. PRIEST 1988: 416). Finalmente, la tercera razón por la cual las contradicciones son especialmente problemáticas en las ciencias es que son indicadores de incoherencias en la comprensión científica global de un fenómeno y por ende obstáculos para las representaciones del mismo.

A favor de esta posición se ha argumentado que si bien los científicos se encuentran de manera recurrente con dificultades que tienen la apariencia de contradicciones, bajo un análisis más profundo de éstas se hace evidente que de hecho no son contradicciones legítimas sino solamente aparentes, y por tanto no se tiene evidencia a favor de la plausibilidad de la tolerancia a la inconsistencia en las ciencias (ver VICKERS 2013: Cap. 8). Para defender esta

tesis se ha ofrecido un argumento de naturaleza epistémica: primero, como tradicionalmente se asume, las contradicciones representan un peligro para la racionalidad; de hecho, si los científicos creyeran una contradicción, esta creencia estaría injustificada y sería irracional. Segundo, la evidencia histórica señala que en la mayoría de los casos, una vez que los científicos descubrían que su teoría era inconsistente, debilitaban sus compromisos doxásticos con, al menos una parte de la contradicción –argumentando que no eran los dos componentes de la contradicción verdaderos sino que al menos uno de ellos era producto de una idealización, de una abstracción o incluso de una ficción. Tercero, como un resultado adicional de este proceso de debilitamiento doxástico, los científicos podían atenuar e incluso diluir la contradicción dado que ya no era un conflicto simplemente entre una proposición s y su negación, sino la conjunción de las proposiciones 's es verdadera' y '¬s es una ficción útil dentro de la teoría' (Cf. VICKERS 2013, DAVEY 2014). Por último, aunque hubiera momentos históricos en los que pareciera que los científicos trabajaban con información inconsistente, dado que no creían la contradicción, no la toleraban; luego, no se tiene evidencia a favor de la tolerancia a la inconsistencia en las ciencias.

La conclusión de este tipo de argumentación es que, no se tiene evidencia de que los científicos hayan tolerado alguna contradicción explícita a nivel doxástico, sin embargo, esto no constituye una respuesta a la pregunta que párrafos arriba describimos como central al debate. Si bien, esta explicación –al menos en [VICKERS 2013]– ha estado acompañada de análisis históricos detallados, dichos análisis se han concentrado en las creencias que los científicos reportaban explícitamente, dejando de lado tanto la estructura formal de teoría como las inferencias que, independientemente de sus compromisos doxásticos, los científicos realizaban con ella. Incluso aceptando la conclusión de este argumento, no es posible desestimar que tanto las teorías científicas como los procesos inferenciales de los científicos emplean en su práctica puedan ser tolerantes a la inconsistencia.

Por otro lado, la *visión tolerante*, en contraste con la visión tradicional, defiende que las teorías científicas pueden y han sido inconsistentes sin por el-

lo poner en peligro la racionalidad de quienes las emplean. De acuerdo con esta visión, la historia de la ciencia contiene suficiente evidencia de que distintas contradicciones han sido toleradas por periodos prolongados, y que durante dichos periodos, los científicos han usado eficientemente para describir, predecir e incluso explicar fenómenos. Es decir, los defensores de la visión tolerante se han ocupado de analizar el fenómeno de tolerancia a las contradicciones en la intersección de los niveles de la teoría y el razonamiento de agentes epistémicos específicos.

Una de las líneas de argumentación más visitadas por la visión tolerante es la siguiente:

a) Al menos, cuando las teorías científicas son jóvenes, es normal reconocer que carecen de algunas de las virtudes epistémicas que teorías más maduras poseen (Cf. MEHEUS 2002).
b) La consistencia no es una virtud privilegiada en la aceptación y rechazo de las teorías científicas; los científicos no están dispuestos a aceptar una teoría solamente porque es consistente, ni están dispuestos a rechazar una teoría que parece inconsistente si es altamente exitosa en dominio de aplicación (Cf. PRIEST 2002, FRISCH 2004).
c) De hecho, la evidencia histórica sugiere que, una vez que se descubre una contradicción en una teoría o modelo empíricamente exitoso, la práctica epistémica más común es tratar de resolver el problema sin abandonar o "archivar" la teoría (Cf. LAUDAN 1977: Cap. 1 y 2) especialmente si se carece de alternativas consistentes que alcancen el mismo éxito empírico.
d) De acuerdo con los registros históricos, existen diversas teorías científicas que, en algún momento se han reconocido como inconsistentes, sin embargo no han sido abandonadas solamente porque sean inconsistentes. Ejemplos de esto son el cálculo infinitesimal (Cf. BROWN y PRIEST 2004, SWEENEY 2014), la electrodinámica clásica (Cf. FRISCH 2004), el modelo del átomo de hidrógeno de Bohr (Cf. BUENO 2006; BROWN y PRIEST 2015), la mecánica newtoniana, el mo-

delo solar estándar (Cf. MARTÍNEZ-ORDAZ 2017, 2020), el modelo del núcleo atómico de capas y el de la gota líquida (MORRISON 2015), así como la teoría de la generación espontánea y la heterogénesis (LAUDAN 1977: Chap. 2).[5]

(🌀) De lo anterior, los defensores de la visión tolerante han concluido que no solamente algunas de nuestras teorías científicas han sido inconsistentes, sino que algunas de las prácticas epistémicas (especialmente, prácticas inferenciales) en las ciencias son tolerantes a las contradicciones.

Ahora, a partir de (🌀), algunos defensores de las lógicas paraconsistentes han realizado una inferencia, aparentemente, abductiva para explicar por qué los científicos pueden trabajar con teorías inconsistentes y no trivializar su razonamiento al mismo tiempo:

e) No existe ninguna evidencia histórica que sugiera que alguna de las teorías enlistadas en (d) fue lógicamente trivial o que, al emplearlas, los científicos infirieron a partir de ellas cualquier proposición de la teoría–tal como PE demandaría que ocurriera.

(🌀🌀) La mejor explicación de porqué (e) es el caso es que el mecanismo inferencial que subyace a la teoría en cuestión o a su aplicación *tuvo que* ser paraconsistente (Cf. MICHAEL 2016: 3355-56).

[5] El hecho de que estas teorías provengan de distintos tipos de disciplinas científicas, desde ciencias formales hasta ciencias de la vida, sugiere que la tolerancia a la inconsistencia no es exclusiva de alguna ciencia en particular, sino que, por el contrario, es un fenómeno que se extiende a través de las diversas áreas de la investigación científica.

Lo anterior puede ser descrito a través del siguiente diagrama:

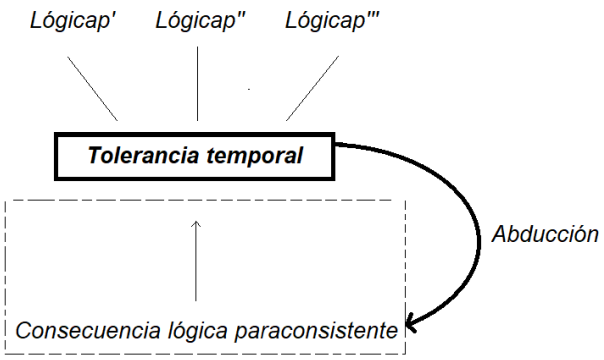


Figura 1. El argumento abductivo de la paraconsistencia

Llamemos a este, el *argumento abductivo (a favor de la paraconsistencia)*. Dicho argumento consiste en, a partir de un caso particular que ilustre tolerancia temporal a la inconsistencia en las ciencias, inferir abductivamente que la mejor explicación de esta tolerancia es que, ya sea el razonamiento de los científicos o la teoría en cuestión, están cerrados bajo una relación de consecuencia lógica paraconsistente (֍֍). Dejando como única tarea a futuro para el lógico o el filósofo de la ciencia, determinar cuál es la lógica paraconsistente específica que permite explicar de mejor el caso particular.

Ahora bien, el mayor problema es que (֍֍) es el resultado de un argumento que "se basa en lo que la gente de hecho infiere o debería inferir, por lo que se basa en un aspecto normativo. La primera pregunta que debe plantearse se refiere al estado del "debería" en la afirmación sobre si deberían haber inferido oraciones aleatorias de la teoría" (MICHAEL 2016: 33556). Esto tiene efectos negativos en dos sentidos: por un lado, debilita la adecuación empírica (histórica) de explicaciones de la tolerancia a las contradicciones –que apelen a la paraconsistencia. Por otro lado, a nivel metodológico, revela un sesgo importante que afecta la legitimidad de la relación entre la evidencia histórica y la aplicabilidad de algunas herramientas paraconsistentes.

Con respecto a la adecuación empírica de (֍֍), algunos filósofos han defendido, contra (֍֍), que existe evidencia importante que apunta a que la racionalidad humana no obedece a principios lógicos, como PE, –que consti-

tuyen teorías de la inferencia particulares. Se ha dicho que los agentes epistémicos responden a retos inferenciales, como las inconsistencias, usando estrategias pragmáticas de razonamiento que ayudan a los agentes a resolver problemas particulares en contextos específicos y que no privilegian teorías de la inferencia determinadas (Cf. HARMAN 1984: 108, MICHAEL 2016: 3355-57). Si esto fuera el caso, complicaría significativamente que (֍֍) fuera verdadera.

A nivel metodológico, esta inferencia abductiva parece injustificada por dos razones. Primero, depende fuertemente de evidencia empírica que nunca es proporcionada –sobre la cerradura lógica de procesos de la cognición humana. Y segundo, asume que todos los casos de tolerancia a la contradicción son, al menos estructuralmente, uniformes, y por tanto, la explicación sobre los mecanismos que permitieron dicha tolerancia, es la misma para cualquiera de ellos.

Incluso concediendo que (֍֍) fue inferida de forma ilegítima, y que la racionalidad humana no está constreñida por alguna teoría de la inferencia específica sino que es independiente de ellas, (֍) permanece aceptable. Aún más importante, el que los proyectos de investigación con mayor rigor historiográfico hayan revelado únicamente que los agentes epistémicos no toleran contradicciones a nivel doxástico, no afecta negativamente la plausibilidad algunas teorías científicas y algunas prácticas inferenciales en las ciencias sean tolerantes a las contradicciones.

Es importante enfatizar que si bien (֍) parece responder satisfactoriamente a la pregunta central del debate sobre la plausibilidad de la tolerancia a la inconsistencia en las ciencias, (֍֍) pertenece a una discusión –aunque relacionada- distinta. (֍֍) constituye una respuesta a la pregunta sobre cómo podemos interpretar, desde un punto de vista lógico, la tolerancia a las contradicciones en la ciencia. Con esto en mente, en la siguiente sección presentaremos presentaremos una estrategia para evitar los problemas metodológicos que tiene este argumento abductivo, y al mismo tiempo permita pasar de (֍) a alguna versión mucho más moderada de (֍֍).

3. Nuestra alternativa

En esta sección, caracterizamos nuestra posición con respecto a la metodología para interpretar, desde un punto de vista lógico, la tolerancia a las contradicciones en la ciencia. Para hacerlo, procedemos en dos pasos: en Sec. 3.1, presentamos algunas aclaraciones pertinentes, en Sec. 3.2, describimos brevemente nuestra posición y en la Sec. 3.3, caracterizamos las distintas contradicciones que es posible identificar en las ciencias empíricas.

3.1. Aclaraciones

Para describir nuestra posición necesitamos hacer algunas especificaciones importantes.

La primera es que, reconocemos que existe un continuo de actitudes epistémicas asociadas a la tolerancia a las contradicciones en las ciencias. Dicho continuo incluye la *aceptación débil* de algunas teorías inconsistentes, en tanto productos epistémicos que vale la pena perseguir con la esperanza de que, a la larga, la contradicción se resuelva y la teoría sea suficientemente virtuosa como para ser considerada como candidata a la verdad (Cf. ŠEŠELJA 2017: 8). De igual forma, consideramos la *aceptación provisional* de Rescher (1988), que permite aceptar una tesis y su negación de manera independiente –pero prohíbe aceptar su conjunción; y la *aceptación restringida* (Cf. DA COSTA Y FRENCH 2003) de acuerdo con la cual, algunas teorías inconsistentes funcionan de forma confiable en dominios de aplicación claramente restringidos (Cf. ŠEŠELJA 2017: 11-13). Por último, se encuentra la *aceptación fuerte*, misma que consiste en aceptar una teoría como candidata a la verdad –incluso a pesar de ser inconsistente.

La segunda especificación es que, en lo que sigue, asumiremos que cualquier análisis de la (in)consistencia de nuestras teorías científicas está constreñido de manera importante por nuestra situación epistémica particular; específicamente por nuestro nivel de conocimiento o ignorancia de la *estructura teórica* de la teoría (Cf. MARTÍNEZ-ORDAZ 2020). La noción de estructura teórica refiere a la estructura inferencial de las teorías empíricas, tal

que expresa patrones inferenciales específicos que garantizan el éxito de la teoría en distintos contextos. Dadas nuestras limitaciones cognitivas, tener acceso a la estructura inferencial completa de una teoría empírica parece en el mejor de los casos, extremadamente difícil, y en el peor, imposible. Sin embargo, en muchas instancias de especialización, los agentes epistémicos adquieren conocimiento sobre fragmentos específicos de la estructura de su teoría, y con ello, la posibilidad de identificar los patrones inferenciales específicos que gobiernan determinados fragmentos de la teoría en cuestión –y así es como pueden determinar cuáles deben ser los productos inferenciales de (fragmentos específicos de) su teoría en contextos particulares.

3.2. Cómo explicar la tolerancia

En lo que sigue, explicaremos brevemente cuál es la causa de (֍) y luego diremos, bajo qué circunstancias, la tolerancia a la inconsistencia –descrita en (֍)– puede privilegiar algunos enfoques paraconsistentes.

Generalmente, la tolerancia a la inconsistencia se ha explicado apelando a cierto tipo de ignorancia: ignorancia de valores de verdad. Primero, existe un supuesto compartido que sugiere que, al menos en las ciencias empíricas, cuando los científicos identifican una contradicción en sus teorías, modelos o fragmentos de razonamiento, suelen asumir que alguno de los dos componentes de la contradicción involucrada es falso (LUDAN 1977: 56).[6] Luego, si los científicos ignoran cuál de esos elementos es falso, pero al mismo tiempo, si no existen alternativas consistentes a la teoría, y ambos elementos de la contradicción son o consecuencias igualmente legítimas de la teoría o supuestos igualmente útiles en la práctica científica, los científicos se encuentran orillados a tolerar la contradicción mientras adquieren el conocimiento que sería necesario para determinar cuál de los componentes de la misma es

[6] ? Sin embargo, el lector puede sentirse inclinado a creer que algunas contradicciones pueden ser verdaderas y por tanto, negarse a aceptar este supuesto, llamemos a esta la "tesis dialetheista". Incluso si concedemos que aceptar la tesis dialetheista puede resultar útil en contextos formales –mayormente asociados a paradojas de autoreferencia-, dentro de las ciencias empíricas, no existe evidencia histórica de que aceptar alguna contradicción como verdadera haya resultado ser la *mejor* alternativa (conceptual o empírica) para resolver un problema científico. Si se desea revisar esta objeción con mayor detalle ver: [MARTÍNEZ-ORDAZ 2020: Sec. 2].

falso (Cf. MARTÍNEZ-ORDAZ 2020). La historia de la ciencia ha revelado que, en la mayoría de los casos, las contradicciones en las ciencias empíricas no son resultado de limitaciones observacionales superficiales, sino que al contrario, son anomalías que permanecen incluso a pesar del desarrollo de instrumentos experimentales que se esperaría, fueran suficientes para resolver el conflicto. El conocimiento del que los científicos carecen en casos de tolerancia a las contradicciones, generalmente, es *conocimiento de la estructura teórica* del fragmento de su teoría que causa la contradicción, lo cual les impide determinar cuál, si alguna, de esas dos proposiciones es falsa.[7] Entonces, no es sino hasta que los científicos exploran ese fragmento de la estructura inferencial de la teoría que pueden determinar si el razonamiento óptimo dentro del mismo es paraconsistente o no.

Lo anterior puede ser descrito a través del siguiente diagrama:

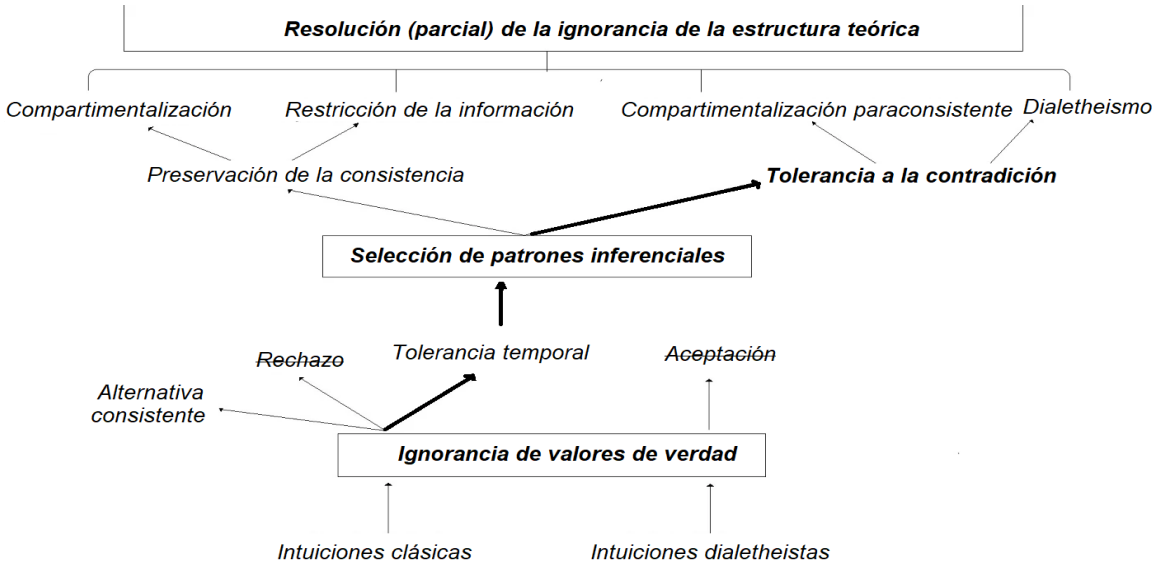


Figura 2. La visión general: ignorancia y tolerancia a la contradicción (en [MARTÍNEZ-ORDAZ 2020: 19]. Nuestra traducción)

> El nivel más bajo del diagrama ilustra que (…) cuando se ignoran los valores de las proposiciones en conflicto, los científicos pueden abandonar el conjunto de información inconsistente y trabajar con una alternativa consistente (si está disponible), o inclinarse a tolerar

[7] La "estructura teórica" de un modelo o una teoría científica empírica es la estructura inferencial que gobierna la teoría o el modelo en cuestión, en la cual se enfatizan patrones inferenciales específicos, de los cales se espera que garanticen el éxito de la teoría en diversos contextos (Cf. MARTÍNEZ-ORDAZ 2020).

> la contradicción temporalmente. patrones de inferencia concretos para trabajar con la información inconsistente, dependiendo del tipo de maniobras inferenciales que eligieron, pueden preservar la consistencia o tolerar la inconsistencia. El nivel más alto muestra que, si se supera (al menos parcialmente) la ignorancia de la estructura teórica, los científicos tienen acceso a la tolerancia a la inconsistencia; y también si se ha logrado la tolerancia a la inconsistencia, esto indica la superación parcial de la ignorancia de la estructura teórica. (MARTÍNEZ-ORDAZ 2020: 19)

Ahora bien, esta explicación no solamente permite reconocer el papel que juega la ignorancia para en la tolerancia a la inconsistencia en las ciencias empíricas, sino que además sugiere una guía metodológica para investigar si los casos individuales de tolerancia a las contradicciones de hecho privilegian alguna herramienta paraconsistente. El camino sugerido consiste en:

- Ir de un caso específico de tolerancia temporal a la selección de posibles patrones inferenciales que permitan trabajar de manera eficiente con el fragmento de la teoría que es inconsistente.
- Hacer una evaluación de los patrones inferenciales alternativos y seleccionar aquellos que permitan a los agentes obtener óptimamente (en términos de recursos computacionales), al menos, los productos inferenciales más valiosos de la teoría.
- Interpretar los patrones inferenciales seleccionados en términos de preservación de la consistencia y de tolerancia a la contradicción (Cf. BUENO 2017), y elegir la interpretación que sea más natural dados los constreñimientos del caso particular –preferencias metodológicas de la disciplina, compromisos doxásticos de los agentes, entre otras.
- Y solamente una vez que se haya hecho toda esta investigación será posible determinar si el caso estudiado, de hecho, constituye evidencia a favor del uso de herramientas paraconsistentes en las ciencias –y cuál, si alguna, es la herramienta que mejor permite reconstruir el razonamiento conducido por los científicos para la tolerancia de la contradicción.

Entonces, para solamente algunos casos históricos que ilustren tolerancia a las contradicciones, los filósofos y lógicos de la ciencia podrán concluir que:

> (֍֍*) La mejor herramienta inferencial de la que disponían los científicos para trabajar satisfactoriamente con su teoría fue/ es paraconsistente.

Esta metodología permite reconocer que hay casos que ilustran tolerancia temporal a una contradicción que no privilegian el uso de herramientas explícitamente paraconsistentes. De igual forma, esta guía metodológica permite hacer análisis más profundos de casos específicos, antes de ofrecer una explicación para el tipo de tolerancia que cada caso ilustra. Además, esta metodología permite ofrecer explicaciones de *cómo* los científicos pudieron razonar con información inconsistente sin poner en riesgo su racionalidad independientemente de compromisos previos con algún tipo de herramientas paraconsistentes –contrario a la metodología que subyace al argumento abductivo (Sec. 2.2). Por último, esta metodología permite identificar como significativas algunas de las diferencias entre los casos particulares, y con ello, reconocer con mayor nitidez y grano fino los patrones inferenciales que, de hecho, jugaron un papel para la tolerancia a las contradicciones en casos específicos. Hacer primero un análisis detallado de los casos específicos, teniendo en cuenta la información de la que disponían los científicos y las formas en las que la ordenaron y la ocuparon; para después ofrecer y comparar interpretaciones alternativas sobre los patrones inferenciales que estuvieron presentes en cada uno de los casos, deja abierta la posibilidad de que incluso casos históricos similares, puedan subyacer patrones inferenciales distintos.

3.3. Distintas contradicciones en las ciencias

La metodología que hemos sugerido para el análisis de tolerancia a las contradicciones y su valor como evidencia a favor del uso de herramientas paraconsistentes en las ciencias, requiere que se hagan análisis detallados del ti-

po de tolerancia temporal que los casos específicos ilustran así como de los patrones inferenciales que permitieron tal tolerancia. Con esto en mente, en esta sección, explicamos cuáles son los distintos tipos de contradicciones que se toleran en las ciencias empíricas para, posteriormente, motivar el estudio de otros tipos de contradicciones en áreas de la investigación científica que han sido tradicionalmente ignoradas.

Primero, en las ciencias es posible reconocer distintos tipos de contradicciones. Cualquier teoría científica (empírica) puede ser inconsistente consigo misma, con otros descubrimientos o con descripciones empíricas que hayan sido fuertemente aceptadas por su disciplina; o bien con otras teorías o modelos de explicación que estén fuertemente aceptadas por la comunidad relevante (KUHN 1977).[8] Es decir,

> si distinguimos entre observación y teoría (lo que no puede ser observado), entonces tres tipos de contradicciones son notables para nuestros propósitos: aquellas entre teoría y observación, aquellas entre teoría y teoría, y aquellas internas a la teoría misma. (PRIEST 2002: 144)

Así, uno puede reconocer, al menos, tres tipos de contradicciones en las ciencias empíricas, es decir una teoría T' puede ser:

- **Internamente inconsistente,** si s y $\neg s$ se encuentran en la subestructura A' de T', y tal que A' contiene explícitamente sólo los elementos teóricos (no-empíricos) de T'.

  Un ejemplo paradigmático de este tipo de contradicciones es la electrodinámica clásica (cf. FRISCH 2004).
- **Inconsistente con la observación,** (es decir, T' sería parcialmente empíricamente inadecuada) si una subestructura empírica de T', A', *que* contiene un enunciado s y, otra subestructura de T', A'', *contiene* $\neg s$.

  Un ejemplo paradigmático de este tipo de contradicciones es

[8] Para el caso especial de las teorías de las ciencias formales uno puede aceptar que es metodológicamente irrelevante determinar si son o no consistentes con descripciones empíricas, o uno puede comprometerse con establecer la aplicabilidad de dichas teorías al mundo empírico como una condición necesaria para el éxito científico.

la anomalía en el corrimiento del perihelio de Mercurio (Cf. PRIEST 2002).

Finalmente, dos teorías T' y T'' serían

- **Mutuamente inconsistentes,** si hubiera una función parcial que asigna elementos de una a la otra y una subestructura T' contiene el enunciado s y una subestructura de T'' contiene el enunciado $\neg s$.
  Un ejemplo paradigmáticos de este tipo de contradicciones es la contradicción entre el modelo del núcleo atómico de capas y el de la gota líquida (Cf. MORRISON 2015: Cap. 5).

Tradicionalmente, el estudio de las contradicciones en las ciencias se ha centrado en las contradicciones internas a las teorías que provienen de disciplinas formales y empíricas (especialmente en distintas áreas teóricas de la física), reduciendo de manera importante el alcance del fenómeno mismo. Por ejemplo, la mayoría de los análisis filosóficos de las contradicciones en la ciencia han concluido que, para el caso de las contradicciones internas, éstas no consisten en un uso irregular del mismo término. El ejemplo paradigmático de esto es el cálculo infinitesimal antiguo: por un lado, la aparente contradicción se encuentra en distintos momentos de una misma prueba, al asumir que el valor del (infinitesimal) ε es *0*, y en otros momentos, asumir que ε es distinto de *0*. Por otro lado, la evidencia histórica ha mostrado que ninguno de los principales proponentes del cálculo antiguo (Newton y Leibniz) consideraron ni que los infinitesimales fueran entidades contradictorias ni que el término en sí mismo se empleara de forma inconsistente (Cf. VICKERS 2013: cap. 6). De hecho, de acuerdo con el registro histórico, si bien pudo haber existido una práctica inferencial inconsistente y no-trivial (cf. BROWN y PRIEST 2004, SWEENEY 2013), ésta práctica no estuvo anclada al uso de un término específico.

Si los análisis de las contradicciones en el razonamiento científico se centran en las ciencias formales y en los fragmentos formales de las ciencias

empíricas, corren el riesgo de ignorar características importantes de la tolerancia a la inconsistencia que tiene lugar en otras áreas del conocimiento científico. Por ejemplo, si uno presta atención a la práctica científica en ciencias sociales, como la economía, uno podría identificar otras presentaciones de las contradicciones y con ello, reconocer formas alternativas de tolerar las contradicciones en las ciencias. Una teoría económica inconsistente puede ser muy distinta de una teoría matemática inconsistente, y los mecanismos de tolerancia y preservación de racionalidad involucrados, posiblemente, serán también distintos.

Por ejemplo, en economía, es posible caracterizar a la teoría del valor de Karl Marx como una teoría internamente inconsistente. En el primer tomo de *El Capital* (MARX, 1991a), Marx asumió que valores y precios (P=V) coinciden para estudiar cómo se determina la tasa de plusvalor en la jornada laboral (la relación entre "tiempo de trabajo socialmente necesario" y "tiempo de trabajo excedentario"). Este mismo supuesto se mantuvo en el tomo II cuando Marx (1992) estudió los esquemas de reproducción del capital a escala social global. Pero ya en el tomo III lo que hizo Marx (1991b) fue levantar el supuesto de que valores y precios coinciden, justo para estudiar otro problema: la formación social de los precios. Algunos críticos de Marx le llamaron a esto "la gran contradicción" (cf. BOHM-BAWER 1949, MANDEL 1992) porque se decía que Marx se estaba contradiciendo a sí mismo. ¿Cómo puede ser posible que en un momento asuma que valores y precios son lo mismo, y en otro momento dice que no son lo mismo? Ante eso, la respuesta está en lo que los marxistas llaman niveles de abstracción (ver MANDEL 1992, NOWAK 1980). En el tomo I Marx está presentando un argumento más general, más "abstracto", sobre cómo funcionan los procesos de valorización y reproducción del capital, y en el tomo III, que realiza un análisis en un nivel más "concreto", donde en realidad está abordando otro problema: la relación entre valores y precios. Particularmente, en las primeras dos secciones del tomo III, Marx se enfocó a estudiar el llamado "problema de la transformación" de los valores (que contienen el tiempo de trabajo socialmente necesario para la producción de las mercancías) a precios corrientes, de mercado. La lección

aquí es extraer la idea de que las "contradicciones internas" de una teoría son dependientes tanto del nivel de abstracción del cual nos estemos refiriendo, como del objetivo para el cual se usan ciertos supuestos. Al igual que en diferentes momentos de una prueba en el cálculo infinitesimal, en el capital de Marx se toleran las contradicciones porque se van cambiando los supuestos en diferentes momentos argumentales. Y esto no necesariamente pone en crisis tanto la racionalidad de los estudiosos de Marx como tampoco el carácter explicativo su obra.

Teniendo este tipo de casos en mente, en las siguientes secciones exploraremos otros tipos de contradicciones en disciplinas científicas distintas de las matemáticas y la física, y las diversas formas en las que son toleradas.

4. Contradicciones y observación

Si el mismo *tipo* de contradicciones puede manifestarse de formas considerablemente distintas en diferentes disciplinas, estudiar y comparar las contradicciones mejor y mayormente toleradas en las diferentes ciencias, podrá sin duda arrojar luz sobre las contradicciones en el razonamiento científico así como sobre los mecanismos más efectivos para su tolerancia. En concordancia con este supuesto, en esta sección nos concentraremos en un tipo de contradicciones que son frecuentemente toleradas en las ciencias empíricas, biológicas e incluso sociales, a saber, las contradicciones entre teoría y observación. Aquí argumentamos que son estas contradicciones, tanto por su frecuencia como por el nivel de tolerancia hacia ellas, las que podrían ser más reveladoras sobre los mecanismos para la preservación de la racionalidad científica en contextos inconsistentes.

Esta sección está dividida en dos partes: en la primera, Sec. 4.1, explicamos qué son las contradicciones entre teoría y observación y defendemos que éstas constituyen el tipo de contradicciones filosóficamente más relevante para el estudio de la tolerancia a la inconsistencia en las ciencias. En la segunda parte, Sec. 4.2, presentamos dos estudios de caso que ilustran tanto la presencia como la tolerancia a este tipo de contradicciones en dos disciplinas científicas distintas –la astronomía y la economía.

4.1. Por qué las contradicciones entre teoría y observación

Las contradicciones entre teoría y observación han sido comúnmente llamadas *anomalías*; sin embargo, dicho término ha abrigado no solamente a estas contradicciones sino también a vacíos tanto explicativos como predictivos (también conocidos como *laguna* kuiperianas). La unificación de estos dos tipos de problemas bajo el mismo término ha sido, sin duda alguna, un grave error que ha tenido como consecuencia que tanto filósofos como lógicos de la ciencia no hayan prestado suficiente atención a las contradicciones entre la teoría y la observación.

Las anomalías consisten en la presencia de un enunciado (generalmente algún tipo de resultado observacional) *s*, tal que al ser combinado con una teoría particular *T*, y con una cláusula ceteris-paribus, el enunciado se convierte en un enunciado potencialmente refutador para la teoría (LAKATOS 1977: 40). No obstante, las anomalías pueden ser de dos diferentes tipos: *lagunas* (KUIPERS 1999, 2000), también asociadas a las *novedades abductivas* (ALISEDA-LLERA, 2006: 47), y *contradicciones lógicas* (Laudan 1977).

Para una teoría (con base empírica) *T*, donde $T = \langle \mathbf{D}; \mathbf{R}_i^n \rangle$, de la cual A_{pred} y A_{exp} son subestructuras empíricas , y *s* es una predicción de *T* y $\neg s$ es un reporte observacional:

> **Una laguna en *T* es** un vacío, ya sea explicativo o predictivo, con la siguiente forma:
>
> 1. $A_{pred} \nvdash s$
> 2. $A_{exp} \nvdash \neg s$
>
> $\therefore$ $\neg s$ es *novedoso* con respecto de *T*
>
> $\neg s$ es novedoso en el sentido de que no está explicado ni predicho por *T*, no obstante es consistente con *T*.

Una contradicción lógica entre *T y* la observación: consiste en contradicciones entre una predicción *s* que se sigue de *T* y una observación ¬*s*. Es decir:

1. $A_{pred} \vDash s$
2. $A_{exp} \vDash \neg s$

$\therefore$ ¬*s* es *novedoso* con respecto de *T*

s es anómalo en el sentido de que no está explicado ni predicho por A_{pred} de *T*.

En otras palabras, una contradicción entre teoría y observación requiere la combinación de lo siguiente:

(i) Dada una teoría empírica T= <D, R_i^n >;
(ii) A_{pred} y A_{exp} son subestructuras empíricas de *T*
(iii) $A_{pred} \vDash s$ y $s \in T$
(iv) $A_{exp} \vDash \neg s$

Ahora bien, la caracterización que aquí hemos presentado de este tipo de contradicciones puede parecer demasiado similar a la ofrecida para las contradicciones internas. De hecho, uno podría creer que ambos tipos poseen estructuras tan similares que uno consiste en el enriquecimiento del otro una vez que un dominio empírico entra en juego. Esta interpretación equivocaría de manera sustancial el punto de atención que aquí nos ocupa.

Al comparar las contradicciones internas con las contradicciones entre teoría y observación, estas últimas resaltan tanto por la naturalidad como por la frecuencia con las que son toleradas. Por un lado, la consistencia interna de las teorías científicas siempre ha sido percibida como un valor epistémico privilegiado; esto principalmente porque se espera que cualquier cuerpo teórico internamente consistente tenga mayor número de consecuencias confiables que uno que sea inconsistente, si la teoría es inicialmente defectuosa en este

sentido, lo más probable es que el problema solamente se agrave una vez que se lleve a un dominio de aplicación específico. Es por esto que, cuando una teoría (ya sea formal o empírica) se sospecha internamente inconsistente, causa una búsqueda inmediata tanto de mecanismos para el control de los daños que la contradicción pueda causar como una investigación que persiga la resolución del conflicto lo antes posible.

Sin embargo, lo anterior no suele ocurrir con las contradicciones entre teoría y observación; por un lado, porque, apoyado en evidencia histórica, de ellas se cree que no tienen el efecto de promover el rechazo inmediato de la teoría en cuestión (Cf. LAUDAN 1977); y por otro lado, porque las contradicciones entre teoría y observación suelen ser atribuidas a la falta de herramientas observacionales con determinado grado de sofisticación –en ese sentido, no son causadas por un problema de la teoría sino por una contingencia instrumental. Si esto fuese cierto, la contradicción podría no tener alguna repercusión importante en la representación de la teoría, como inconsistente, pero, sin duda, la tendría a nivel del razonamiento conducido por los científicos al manejar dicho conjunto de información inconsistente. Dado que las contradicciones entre teoría y observación son tanto las más comunes como las mejor toleradas en las ciencias, su estudio detallado permitirá alcanzar una mejor comprensión de la tolerancia a la inconsistencia en las ciencias.

Es bien sabido que las ciencias empíricas legitiman, a través de sus respectivas metodologías, el papel de la observación; por tanto, cuando una contradicción entre la teoría y la observación aparece, este conflicto afecta fuertemente la base de la aplicabilidad de la teoría y no debe ser confundido con una nimiedad –la teoría está fallando al cumplir uno de sus principales propósitos. A pesar de esto, los filósofos y científicos estarían de acuerdo en que al identificar una contradicción de este tipo ni la teoría en cuestión ni los reportes observacionales deben ser irremediablemente abandonados. A este respecto, se ha sostenido que en la mayoría de los casos, tanto la teoría como los reportes pueden aceptarse *pro tem* mientras se elabora una nueva teoría o se desarrollan nuevos instrumentos. Algunos ejemplos de este tipo de contradic-

ciones son la anomalía en la precesión del perihelio de Mercurio y la hipótesis de Prout (cf. LAUDAN 1977: Cap. 2; PRIEST 2002).

A la luz de lo anterior, si se desea analizar las contradicciones en las distintas disciplinas científicas que tengan consecuencias observacionales, los aspectos ligados a la observación no deben quedar al margen de ningún modo. Habiendo dicho esto, en los siguientes párrafos, presentamos brevemente dos estudios de caso que ilustran contradicciones entre teoría y observación en dos disciplinas científicas distintas.

4.2. Ejemplares

El ejemplar mejor aceptado de este tipo de contradicciones es la anomalía en la precesión del perihelio de Mercurio. En pocas palabras, de acuerdo con las leyes de Kepler y con la teoría gravitacional de Newton, todos los planetas orbitan alrededor del sol siguiendo una trayectoria elíptica fija. Sin embargo, en 1859, Le Verrier descubrió que cuando la órbita de Mercurio terminaba, no regresaba al mismo punto al final de cada órbita. El problema radicaba en el hecho de que aunque todos los planetas presentan una precesión en su perihelio, el caso de Mercurio rebasaba el grado de esta precesión. En 1859, Le Verrier anunció que la diferencia entre reportes predictivos y observaciones en la órbita de Mercurio era de 38 segundos sexagesimales por siglo (HARPER 2007: 937). De acuerdo con las leyes de Newton, la elipse de su órbita debía tener una precesión de 432 segundos sexagesimales por siglo, pero en las observaciones notó que la tasa era de 472 segundos sexagesimales por siglo; en términos generales, la teoría relevante no predecía una precesión y tampoco podía explicar los movimientos en la órbita de Mercurio.

Este caso suele aceptarse como una instancia de tolerancia a una contradicción entre teoría y observación apelando a la conjunción de cuatro factores importantes. Primero, la diferencia entre la predicción y el reporte observacional era significativamente más grande que el margen de error aceptado en aquel momento histórico, lo cual permitía interpretar a una como la negación de la otra. Segundo, la órbita de Mercurio era un fenómeno que pertenecía legítimamente al dominio de aplicación de la teoría, y de hecho, al

menos en un sentido teórico, la teoría predecía su comportamiento. Tercero, a pesar de los distintos intentos por corregir la predicción para que fuera empíricamente adecuada, esto no había sido posible; la teoría, en su estado de mayor sofisticación matemática, seguía arrojando la misma predicción. Finalmente, el fracaso de los sistemáticos intentos por ya fuera explicar o corregir el resultado observacional legitimaron el reporte observacional que contradecía la predicción de la teoría. Si bien los científicos de la época no consideraban como candidatos a la verdad a la predicción de la teoría y al reporte observacional, tampoco poseían las herramientas teóricas o metodológicas necesarias para desestimar alguno de ellos. Lo cual los dejaba en una situación donde aceptaban fuertemente la teoría newtoniana y las leyes de Kepler, así como el supuesto de que no había nada en Mercurio que le hiciera esencialmente distinto del resto de los planetas que la teoría predecía de forma adecuada; esto impedía que los científicos pudieran desestimar la predicción –sin que dicho movimiento fuera ad hoc (ver MARTÍNEZ-ORDAZ 2017). A la par, la fuerte confianza que se tenía en los instrumentos observacionales, misma que robustecía el reporte observacional, impedía el rechazo racional de dicho reporte.

La combinación de estos factores hizo que el conflicto entre la teoría y el reporte observacional se reconociera como una contradicción, no era posible desestimar la predicción –pues estaba fuertemente respaldada por el éxito de la teoría con respecto a otros planetas; además, no era posible desestimar a Mercurio como un modelo esperado de la teoría, lo cual justificó la conjunción de la predicción y el reporte. Además, la diferencia entre la observación y la predicción era lo suficientemente grande como para aceptar que una era interpretable como la negación de la otra. Finalmente, el conflicto se mantuvo por casi un siglo hasta que apareciera la teoría de la relatividad general de Einstein. El largo tiempo durante el cual los científicos pudieron emplear la teoría aun conociendo la contradicción que existía con la observación, es evidencia a favor de la tolerancia a las contradicciones en este caso.

Un ejemplo análogo que ilustra la presencia de contradicciones entre teoría y observación en las ciencias sociales es lo que en macroeconomía se

conoce como estanflación. Este fenómeno es aquel que se produjo durante la década de los setenta del siglo XX cuando se observaron algunos casos de países donde hubo creciente inflación y estancamiento económico (lo cual generaba desempleo). Esta situación no era clásica, pues los economistas usualmente habían esperado tendencias que muestren una correlación entre tasas de inflación y tasas de desempleo, donde hay un *intercambio* (en inglés, *trade-off*) entre desempleo e inflación (la conocida "curva de Phillips"): a mayor desempleo - menor inflación, a menor desempleo - mayor inflación.

Durante el periodo de la posguerra, a partir de 1945 hasta 1970, las posturas neokeynesianas dominaron el terreno de las políticas económicas basándose en la curva de Phillips. Ellos planteaban una elección para los gobernantes: o mantener baja la inflación o disminuir el desempleo (cf. HOOVER 2008). Lo primero (tener un nivel general de precios constante) implicaría relentizar la economía y con ello evitar que haya disminución del desempleo. Lo segundo (tener un menor desempleo) implicaría aumentar la actividad económica y con ello aumentar la inflación. El gobierno podría, desde el punto de vista de las políticas económicas neokeynesianas, tener políticas fiscales y monetarias expasivas para así aumentar el crecimiento económico y con ello conseguir el pleno empleo (una situación en la cual todos los trabajadores que quieran trabajar consigan un empleo), pero ello implicaría tener niveles de inflación más o menos altos (cf. HOOVER 2008, GORDON 2011). Sin embargo, la observación del fenómeno de la estanflación contradecía las situaciones planteadas en los modelos neokeynesianos, y llevó a que los economistas tuvieran que repensar sus modelos macroeconómicos. Tuvieron que aceptar que hay casos en los cuales las políticas fiscales y monetarias expansivas no podrán aumentar el empleo, y con ello, se llevó a replantear la curva de Phillips. Milton Friedman (1968) y Edmund Phelps (1967) la reinterpretaron señalando que esta curva podría adoptar diferentes formas: podría tener una pendiente negativa en el corto plazo, pero sería completamente vertical en el largo plazo –donde el incremento de la inflación no reduciría el desempleo. Pero en cualquier caso, la curva de Phillips clásica (planteada como el intercambio infla-

ción-desempleo) no debería verse como un caso general, sino particular (cf. GORDON 2011).

Aquí el punto es ver cómo interpretar la contradicción teoría-observación, y ver las implicaciones que tiene esa contradicción para reformular las teorías. En economía se ha buscado construir enfoques que luego unifiquen los casos particulares de contradicción teoría-observación (como en física). Pero parece que no hay consensos sobre qué teoría tendría que ser la base para establecer dicha unificación, más aún porque en economía hay una gran variedad de escuelas de pensamiento económico. En el tema macroeconómico, por lo menos destacan los post-keynesianos, los neokeynesianos, los monetaristas, los teóricos de las expectativas racionales que su vez se dividen en dos (los nuevos clásicos y los nuevos keynesianos), así como los nuevos post-keynesianos que buscan microfundamentar la macroeconomía en la economía del comportamiento.

A pesar de que estos episodios históricos son considerados como casos paradigmáticos de contradicciones en las ciencias, en la siguiente sección defenderemos que estos casos ilustran una de las instancias más sencillas de las contradicciones entre teoría y observación, y que los casos filosóficamente más interesantes, son mucho más complejos que los dos que presentamos en esta sección.

5. Tres subtipos de contradicciones entre teoría y observación

Hemos defendido que el estudio de las contradicciones entre teoría y observación permite centrar la atención en los casos no solamente más comunes sino mejor tolerados de contradicciones en las ciencias. En esta sección abordaremos la pregunta de si son todas las contradicciones entre teoría y observación del mismo tipo; a saber, si a nivel estructural se comportan de manera uniforme a través de las distintas disciplinas científicas. De ser la respuesta una negativa, el estudio de cada uno de los subtipos de contradicciones entre teoría y observación, a la larga, permitirá el reconocimiento de diferentes mecanismos efectivos para la tolerancia a las contradicciones.

Esta sección está dividida en tres partes: en la Sec. 5.1, discutimos las similitudes que existen entre la anomalía en el corrimiento del perihelio de Mercurio y el caso de los modelos econométricos keynesianos. Ahí mismo, presentamos la conclusión más reveladora que se ha obtenido del análisis de casos históricos que ilustran tolerancia a las contradicciones en las ciencias, y el impacto que dicha conclusión ha tenido en la detección de mecanismos (mayormente inferenciales) que expliquen dicha tolerancia. En la Sec. 5.2, nos concentramos en presentar un subtipo de contradicciones entre teoría y observación cuya detección requiere incorporar un marco teórico adicional a la teoría y al reporte observacional. En la Sec. 5.3, consideramos las contradicciones entre teoría y observación que emergen cuando no existe independencia entre la teoría que genera la predicción y el reporte observacional. Ilustramos cada uno de estos subtipos con estudios de caso de economía.

5.1. Contradicciones con independencia observacional

La conclusión que tanto lógicos y filósofos de la ciencia han extraído de la anomalía del corrimiento del perihelio de Mercurio ha sido que estas contradicciones no son problemáticas para la ciencia porque es muy sencillo aislarlas mientras se mantiene en uso la teoría. Por ejemplo, en [DAVEY 2014] se argumenta que una vez que los científicos se encuentran con una contradicción en su teoría o modelo, dejan de estar justificados en creer en la parte de la teoría que está asociada a la contradicción. En particular, cuando la predicción de una teoría falla, los científicos dejan de confiar tanto en el subconjunto de enunciados involucrados en la inferencia que lleva a la predicción como en el subconjunto de enunciados involucrado en la construcción del reporte observacional, de modo que la parte de la teoría que permanece consistente es la que se mantendrá en uso. Por esta razón, incluso si concedemos que la anomalía en el corrimiento del perihelio de Mercurio es una contradicción, es fácil explicar el razonamiento que permitió que los científicos siguieran trabajando con la teoría sin tener que trabajar con un conjunto inconsistente de información.

Esta explicación ha sido bien recibida por filósofos y lógicos de la ciencia, y ha jugado un papel central en el escrutinio del razonamiento inconsistente no-trivial. De hecho, la estrategia de división (e incluso, segregación) de información se ha convertido en la metodología más común para explicar la forma en la que los científicos pueden trabajar con información inconsistente sin poner en riesgo su racionalidad. Las teorías pueden ser, en abstracto, globalmente inconsistentes, pero los segmentos de las teorías con los que los científicos de hecho trabajan son internamente consistentes (cf. BROWN y PRIEST 2004, 2015).

La estrategia de división arriba mencionada es posible en gran parte porque estos casos de contradicciones entre teoría y observación involucran *exclusivamente* una predicción de una teoría dada y un reporte observacional específico. Si una contradicción entre teoría y observación involucra exclusivamente una predicción y un reporte observacional, el siguiente criterio se satisface:

> **Criterio de independencia observacional:** el conjunto de proposiciones que subyacen al diseño de instrumentos y métodos utilizados para evaluar las consecuencias observacionales de *T*, idealmente se logran con independencia total de las proposiciones pertenecientes a la teoría en cuestión (MARTÍNEZ-ORDAZ, 2017: 140).

En lo que sigue, llamaremos a los casos en los que este criterio se cumple *contradicciones entre teoría y observación con independencia observacional*, y los definiremos como sigue:

> **Contradicción T-O (Indp):** Dada una teoría empírica *T* que tenga a *s* como una consecuencia observacional, si un experimento se realiza, ¬*s* se reporta: igualmente, hay una intersección vacía entre los subconjuntos de los supuestos relevantes de *T* y los supuestos relevantes de la teoría detrás del diseño de los instrumentos y del experimento.

Algunos ejemplos provienen de la economía experimental y del comportamiento, donde se han realizado experimentos que han sometido a prueba empírica algunas predicciones que se siguen de los modelos estándar de elec-

ción racional. Los modelos estándar enfatizan en el carácter coherente de la racionalidad. Una persona es racional en tanto toma decisiones que son coherentes, tanto con sus preferencias como con el paso del tiempo. Esto porque se asume que el agente cumple con dos axiomas básicos: completitud y transitividad. El primero supone que los agentes tienen una lista de preferencias perfectamente ordenadas que les permite decidir entre dos o más opciones. El segundo supone que tienen ordenamiento de preferencias consistentes, donde si se prefiere A sobre B, y B sobre C, entonces, si se es racional, se tiene que preferir A sobre C. Sin embargo, en los experimentos se ha documentado una inconsistencia entre los que los agentes prefieren y lo que eligen (cf. THALER 1980, 2000). Por ejemplo, en *t1* se pudo haber preferido A sobre B, en *t2* preferir B sobre C, pero en *t3* preferir C sobre A. Este ejemplo refuta la tesis de Samuelson de que las preferencias son "reveladas" por las elecciones observadas por los agentes. Es decir, las elecciones de los agentes observadas no necesariamente nos revelan lo que los agentes prefieren debido a que existe un problema de autocontrol, donde la toma de decisiones puede darse por factores impulsivos, o emocionales (cf. THALER 1980, 2000). Este problema de falta de autocontrol lleva a que las personas suelan tener un comportamiento inercial, donde toman decisiones sin darse cuenta, mostrando así una incoherencia intertemporal en los comportamientos de los agentes (cf. THALER 1980, 2000).

Lo mismo aplica para el caso de "sesgos de estado presente" (cf. O´DONOGHUE y RABIN 1999) donde las personas pueden elegir opciones que reportan una utilidad esperada menor a la que podrían haber logrado obtener con otras opciones que se abren en el futuro (por ejemplo, usar el dinero para consumir algo que reporta una utilidad inmediata, y no ahorrar para lograr algo que puede reportar más utilidad futura). En estos casos se tienen comportamientos no optimizadores que se alejan de las predicciones hechas por los modelos estándar de elección racional (ver TVERSKY y KAHNEMAN 1974). No por ello el modelo estándar debería abandonarse. Pero su carácter descriptivo se tiene que repensar.

La característica más destacada de este tipo de contradicciones es que es posible distinguir y separar las predicciones de la teoría de los reportes de observación. El estudio exhaustivo de estas contradicciones y de sus propiedades más generales le ha permitido a lógicos y filósofos de la ciencia concluir que la estrategia de división mencionada en párrafos anteriores, es la forma más eficiente de trabajar con contradicciones en la ciencia.

En la literatura, se reconocen al menos tres formas en las cuales los científicos pueden adoptar esta metodología de división:

- **Aislamiento:** Cuando una predicción y un reporte observacional se contradicen mutuamente, bajo el supuesto de que al menos uno de ellos es falso, los científicos dejan de usar ambos y continúan empleando y confiando tanto en el resto de la teoría como en resultados observacionales similares (ver DAVEY 2014). La desventaja más clara de seguir esta metodología es que tanto la predicción como el reporte observacional dejan de ser accesibles en la práctica científica –y con ello, existe una reducción del dominio de aplicación de la teoría.
- **Debilitamiento doxástico:** Cuando se tienen una predicción *s* y un reporte observacional ¬*s*, los científicos pueden seguir trabajando con ambos si y solo si (a) disminuyen su grado de compromiso doxástico con al menos uno de los elementos de la contradicción, y (b) tienen una explicación que justifique tal debilitamiento. Por ejemplo, si un modelo altamente idealizado predice *s*, pero en la observación se reporta ¬*s,* el agente puede apelar a la naturaleza de dicho modelo para justificarse en no creer que *s* pueda ser satisfecho en el dominio empírico (para una defensa de este tipo de estrategia, cf. VICKERS 2014: Cap 8). La desventaja más importante de seguir esta metodología es que deja de lado los procesos inferenciales que los científicos, independientemente de sus preferencias doxásticas, siguen al trabajar en contextos inconsistentes.
- **Fragmentación (y permeabilidad):** Si se tiene un conjunto inconsistente de información, la forma más eficiente de trabajar con él es fragmentarlo en subconjuntos lógicamente consistentes y restringir el flujo de la información de un subconjunto a otro –evitando, al menos, la formación de contradicciones en dichos subconjuntos (cf. BROWN y PRIEST 2004, 2015).

> Las principales ventajas de esta estrategia son que es altamente compatible con (2) y que además le permite a los agentes usar toda la información de la teoría y la observación –es decir, en contraste con (1), aquí no es necesario renunciar a alguna parte del conjunto inconsistente de datos. Además, la práctica de fragmentar conjuntos inconsistentes en subconjuntos consistentes y permitir que solo una cantidad reducida de información salga de un subconjunto para penetrar en otro.

Ahora bien, queremos enfatizar que, independientemente de cuál de estas estrategias de división, si alguna, pueda ser más útil al trabajar en contextos inconsistentes, el reconocimiento y la formulación de las mismas han sido consecuencia directa de un análisis detallado de un tipo específico de contradicciones. Por tal razón, de haber otros tipos de contradicciones entre teoría y observación, su estudio sería, sin duda alguna, revelador con respecto al tipo de estrategias que pueden ser usadas para su tolerancia.

5.2. Contradicciones desde un marco adicional

En las ciencias, las diferentes disciplinas y dominios de investigación nunca son completamente independientes entre sí (cf. LAUDAN 1977: 53; ELSAMAHI 2005: 335). De hecho, considerando que ciertos problemas científicos son suficientemente complejos como para resolverse usando solamente una teoría o un solo modelo; con más frecuencia de lo que podría pensarse, los científicos combinan distintas teorías (de diferentes disciplinas) y distintos modelos (de diferentes teorías) para resolver problemas específicos. Dado que los científicos tienden a agrupar distintas teorías y modelos para extender su alcance lo suficiente como para explicar fenómenos complejos, los conjuntos de información que se forman por estas combinaciones serían lo que aquí llamamos *marcos de trabajo científicos* (en adelante, *marcos de trabajo*).

Un marco de trabajo consiste en una combinación de conjuntos (originalmente) distintos de información que se relacionan de tal modo que un subconjunto que contiene todos los supuestos fundamentales de cada teoría o modelo se encuentra combinado de un modo, al menos, no-trivial, lógica-

mente consistente mutuamente compatible y (débilmente) complementario. Finalmente, los elementos de un marco de trabajo particular están sujetos a "nuevas" reglas inferenciales; no necesariamente compartidas por todos los conjuntos de información "originalmente" diferentes.

Una particularidad de casos como la anomalía en el corrimiento del perihelio de Mercurio es que la falta de teorías alternativas combinada con la robustez de los reportes observacionales, hacen que el reconocimiento de anomalía como una contradicción sea sencillo –e incluso intuitivo para la comunidad científica. Sin embargo, creemos que la nitidez propia de este estudio de caso es más una excepción que una norma. Para la gran mayoría de casos que ilustran tolerancia a la contradicción en las ciencias, es necesario incorporar elementos teóricos adicionales para poder distinguir si la anomalía es de hecho una contradicción –o si es el resultado de una laguna predictiva o explicativa.

Consideremos lo siguiente: si una observación *r* no fuera explicable apelando a un margen de error relevante, esto no garantiza que *r* contradiga a una predicción *s*. De hecho, *r* podría ser el resultado de un vacío, explicativo o predictivo, entre predicción y observación. Si la anomalía fuera causada por una ausencia de información suficiente, no sería sorprendente que los científicos no pudieran predecir todos los fenómenos que se espera predigan. En esta clase de casos, es posible afirmar haber identificado una contradicción entre teoría y observación si y sólo si existe un marco de trabajo, distinto de la teoría que está siendo evaluada, tal que muestra que, en última instancia, la información disponible para los científicos es suficiente para proveer predicciones adecuadas acerca de *D*, pero la teoría provee una predicción que no se satisface; y los reportes experimentales muestran que la diferencia entre la predicción y la observación es más grande que el margen de error relevante.

En lo que sigue, llamaremos a este subtipo de contradicciones *contradicciones entre teoría y observación a través de un marco de trabajo adicional* (contradicciones T-O+M) y las caracterizaremos de la siguiente manera:

> **Contradicción T-O+M:** Dada una teoría empírica *T* que tiene *s* como consecuencia observacional, si se realiza un experimento se reporta *r*. Pero hay otro marco teórico, *M*, que provee evidencia adicional que legitima interpretar a *r* como ¬*s*.

Aquí también, ejemplos de economía del comportamiento y economía experimental donde se muestra que un comportamiento no esperado (beta en vez de alfa), pero que puede llevarnos a interpretar beta como no-alfa. Por ejemplo, la economía estándar asume que los agentes económicos típicos (los llamados *homo economicus*) basan su comportamiento en dos elementos clave: el interés propio y la racionalidad perfecta (cf. ELSTER 2012). El interés propio se entiende como una motivación básica que lleva a las personas a elegir cursos de acción que satisfacen intereses materiales personales. Los modelos estándar suelen asumir que el motivo único, o principal al cuales todas las otras motivaciones (como el altruismo o el compromiso con normas sociales) se reducen es el interés propio. Este supuesto lleva a asumir que el agente busca maximizar su utilidad propia eligiendo opciones que recompensan su bienestar y evitando opciones que minimizan su utilidad –como castigos y costos. El supuesto de que los agentes tienen racionalidad perfecta se complementa con el interés propio. Este se entiende como una capacidad de cálculo que ayuda a los agentes a elegir entre diferentes medios para lograr fines ya preestablecidos. Es un tipo de racionalidad instrumental. La idea central de los modelos estándar de elección racional es que se asume que los agentes realizan un tipo de cálculo egoísta que busca maximizar sus funciones de utilidad a puntos óptimos, eligiendo exactamente la opción, la cantidad y el momento exacto que maximiza la utilidad. Eso se logra por medio del cálculo de derivadas parciales. En un primero momento se aplicaron estos modelos sólo a temas económicos (como empresas que deben elegir entre cantidad y costos de producción y consumidores, que deben elegir entre cantidades y precios), pero luego se aplicaron a otros casos, como elección de pareja, criminología y corrupción en un proceso que se llamó "universalización del modelo de utilidad esperada" (ver BECKER 1992, LAZEAR 2000). Es importante decir que es-

tos modelos asumen que los agentes actúan como si fuesen perfectamente racionales, que calculan probabilidades siguiendo el teorema de Bayes y que su motivo único para elegir es el interés propio (ver FRIEDMAN 1953). Pero eso no significa que en esos modelos se comprometan con el realismo de sus supuestos. Simplemente se realizan esos supuestos para poder simplificar y con ello facilitar la práctica de la modelización.

En el caso de la corrupción, en los modelos estándar se asume que el agente puede actuar de manera corrupta (transgredir normas legales) si el acto corrupto es rentable. Es decir, la utilidad de ganar más haciendo el acto corrupto es mayor que no hacer corrupción y seguir las reglas legales y morales que hay en una situación dada. Este carácter rentable, a su vez, se multiplica por la probabilidad de ser detectado y de ser castigado (cf. ROSE-ACKERMAN 1974). Se trata de la aplicación del modelo de utilidad esperada para casos de riesgo e incertidumbre. En estos casos, se asume que todos los agentes que se comportan como los *homo economicus*. La predicción es que todas las personas actuarían de manera corrupta, maximizando su interés propio individual si se les presentara la oportunidad. Es decir, cuando transgredir las normas sea rentable y que la probabilidad de ser castigado sea baja (cf. ROSE-ACKERMAN, 1974). Para disminuir la corrupción, se tendría que diseñar un ambiente institucional donde ser corrupto no sea rentable y que la probabilidad de ser castigado por transgredir las normas sea alta.

Sin embargo, este modelo ha sido cuestionado por la economía del comportamiento, la cual ha buscado des-idealizar la noción de agente racional para sustituirlo por supuestos más realistas (cf. HEIDL 2016). Y la manera de hacerlo es, usualmente, construyendo datos empíricos por medio de experimentos que reproducen situaciones donde los agentes interactúan bajo condiciones controladas las cuales nos permiten comparar cómo actuarían los *homo economicus* y como de hecho actúan las personas de carne y hueso. Para seguir con el caso de la corrupción, Dan Ariely (2008), un economista del comportamiento, sometió a prueba empírica la hipótesis estándar que consiste en la afirmación de que todas las personas buscarían transgredir las normas si el-

lo es rentable, se les presenta la oportunidad y que lo harían aprovechando plenamente la ocasión para optimizar sus funciones de utilidad.

Mazar, Amir y Ariely (2008) diseñaron y aplicaron un experimento donde han estudiado si los estudiantes de negocios de escuelas como Harvard, MIT, UCLA entre otras se comportarían de manera deshonesta si tuvieran la oportunidad, y si es así, qué tanto la aprovecharían. En el experimento se tenían tres grupos de estudiantes a los cuales les dieron unos cuestionarios para que estos respondieran. Por cada respuesta correcta se les daba una recompensa (10 céntimos). En un primer grupo los estudiantes contestaron correctamente 32% en promedio de las preguntas. A esos estudiantes simplemente se les recogió el cuestionario y se les evaluó. No tenían la oportunidad de "engañar". Este fue el grupo de control. En un segundo grupo se introdujo una variación: los estudiantes se autoevaluaron. Después de terminar el cuestionario se les dio una hoja que contenía las respuestas correctas. Los estudiantes podían ver las respuestas correctas. Se detectó que en este grupo subió el promedio de las preguntas que se contestaron de manera correcta, a 36%. Esto significa que, o bien este grupo tuvo un mejor desempeño que el otro, o que hubo algunos estudiantes que hicieron trampa. Cuando se revisaron los cuestionarios se vió que había más borrones que en el cuestionario anterior, por lo que es posible decir que más estudiantes hicieron trampa. En el tercer grupo se introdujo otra variación: el evaluador ya no estuvo presente al momento en que los estudiantes se autoevaluaron. Lo que se detectó es que, igual, el promedio de las preguntas correctamente contestadas fue de 36% aproximadamente. Este tercer grupo, sin embargo, tuvo más oportunidad de actuar de manera más tramposa que el segundo, pero los resultados muestran que "no aprovecharon la oportunidad" que, al menos, un *homo economicus* si aprovecharía dado que los riesgos de ser detectado mientras se hace trampa eran más bajos.

En este experimento, Dan Ariely (2008) señala que los estudiantes, si bien pueden decir que ellos mismos son gente honesta (así se perciben a sí mismos), si se les presenta una oportunidad para actuar de manera deshonesta para sacar provecho, aunque sea de manera aislada, lo harán. Sin embargo,

hay algo más: los estudiantes que pudieron haber engañado más (el tercer grupo) para sacar más provecho (nadie los vigilaba), no lo hicieron. Es decir, tuvieron un comportamiento que no fue maximizar su utilidad. Esto, según Ariely (2008), se explica por el hecho de que los agentes tienen ciertas normas sociales y valores que les hace moderar su comportamiento deshonesto. Su deshonestidad fue acotada. Este tipo de hechos sugiere, como otros teóricos lo han dicho, que la racionalidad de los agentes también está éticamente acotada (ARIELY, 2008).

En otro experimento se les dijo a los alumnos, que antes de contestar las preguntas, escribieran los diez mandamientos de la biblia, lo que ellos recordaran. La mayoría no los recordó de manera completa. Sin embargo parece que este ejercicio de recordar tuvo una influencia en el comportamiento del grupo porque, al momento de autocalificarse, estos tuvieron en promedio 32% de preguntas correctamente contestadas. Eso apesar de que los alumnos se autoevaluaron y el evaluador salió del lugar. Ariely señala que en este caso parece que la manera en cómo se presentó la información, donde se introdujo el ejercicio de recordar un conjunto de normas morales, tuvo un efecto: inhibe el comportamiento deshonesto de algunas personas cuando se les recuerdan las normas morales. Quizá esto se debe a que los agentes, como ya lo han dicho Kahneman y Thaler, también toman decisiones basados en algún tipo de emoción o sentimiento de justicia que no necesariamente se pueden verbalizar, pero que están ahí en los agentes. El tema de las emociones y el sentido de justicia (*fairness*) ha sido ampliamente estudiado por los economistas del comportamiento donde se muestra que este sentimiento puede jugar un papel activo en relaciones de reciprocidad.

Los experimentos de Dan Ariely nos recuerdan la importancia de la "racionalidad acotada o limitada" (*bounded rationality*). Por "racionalidad acotada" nos referimos a agentes que toman decisiones que no son óptimas ni que maximizan su utilidad. Al contrario, los agentes tienen una serie de sesgos e ilusiones cognitivas que les impide procesar toda la información de la mejor manera posible, los lleva a cometer errores severos y sistemáticos por lo que su capacidad de aprendizaje es limitado. Pero además, la racionalidad es ética-

mente acotada, en el sentido de que los agentes no actuan sólo con base en el interés propio, sino que actúan con base en el compromiso con ciertas normas sociales (y morales) que acota el actuar del interés propio. Y en unas ocasiones, directamente sustituyéndolo por preferencias pro-sociales (cf. BOWLES y POLANIA-REYES 2012).

En los enfoques de economía del comportamiento, la corrupción se enfoca como un problema de toma de decisiones, pero que tiene una serie de factores que vuelven más complejo entender las motivaciones reales de los agentes, no reducibles al mero interés propio y a los incentivos. Este enfoque apunta a tomar más en serio factores psicológicos. Con lo cual la tendencia hacia la des-idealización del agente racional se ha reforzado en algunas escuelas de pensamiento económico.

5.3. Contradicciones entre una teoría, una teoría auxiliar y la observación

En lo que sigue consideramos los casos en los que el Criterio de independencia observacional no se satisface; en particular, en los cuales esto ocurre por un grado significativo de holismo entre la teoría que está siendo evaluada y ya sea la teoría que subyace al reporte observacional, o alguna otra teoría auxiliar.

Nos ocupamos de este tipo de casos porque consideramos que, frecuentemente, cuando se identifica una predicción que contradice un reporte empírico, el problema no radica sólo en las proposiciones que explícitamente se saben en conflicto. Comúnmente, tales proposiciones mutuamente inconsistentes son dependientes de muchas otras partes de la teoría que no siempre son fáciles de identificar, y en las cuales los científicos confían plenamente (a pesar de su contribución a la ahora reconocida contradicción). Estas contradicciones están gobernadas por el holismo de las relaciones entre la teoría que está siendo evaluada, la observación y la teoría que subyace al reporte observacional.

Cuando nos referimos al holismo entre las teorías hay dos aspectos principales que deben considerarse: las relaciones holistas internas y las externas. Por una parte,

> "[L]as descripciones teóricas de cualquier sistema (físico, económico, etc.) generalmente involucran un complejo extenso de hipótesis. Este complejo incluye principios básicos de una o más teorías empíricas, leyes especiales, hipótesis auxiliares, condiciones de marco, etc. A pesar de que el surgimiento de un conflicto con los datos disponibles significa que este complejo ha fallado como un todo, no es completamente claro cuál de los componentes es fallido y debe ser modificado" (GÄHDE 2002: 69-70).

Por otra parte, en la ciencia las diferentes disciplinas y dominios de investigación nunca son completamente independientes uno del otro. Entonces, es preciso reconocer que el holismo no sólo afecta a los elementos de una teoría aislada; pues tal cosa como un conjunto aislado de creencia parecería más la excepción que la regla en la ciencia:

> [L]a descripción teórica de un sistema rara vez se da aisladamente, sino correlacionada con la descripción teórica de otros sistemas en diversos modos (…) Tales correlaciones son de mayor consecuencia en caso de que haya una discrepancia entre teoría y observación. Si dicho conflicto surge, las modificaciones no necesariamente deben-comenzar en la descripción teórica del sistema en el que el conflicto fue observado. Más bien los esfuerzos correctivos deben comenzar-con el tratamiento teórico de otro sistema correlacionado con el primero (GÄHDE 2002: 69, 70).

Esto significa, en términos generales, que cuando un problema tal como la presencia de contradicciones se descubre en una teoría, cualquier modificación llevada a cabo para arreglar la teoría —ajustes como el aislamiento de inconsistencias— requerirá indirectamente de modificar no sólo otras partes de la teoría misma sino también partes de otros cuerpos de conocimiento relacionados, e incluso así podría no tenerse seguridad de que la contradicción ha sido 'removida.' De modo que si bien hablar de detección de inconsistencia apelando al holismo que a veces está presente en las teorías científicas, debe decirse que a veces el grado de holismo de una teoría impide el uso exitoso de alguna estrategia de división, y que si se ha de mantener la teoría en uso es necesario tolerar la contradicción aunque sea de manera temporal.

Ahora bien, incluso si el lector está de acuerdo en que estas contradicciones se encuentran presentes en la práctica científica, puede defender qwue no son contradicciones entre teoría y observación, sino, contradicciones entre dos teorías – por ejemplo , la que está siendo evaluada y la que subyace al diseño

experimental. Sin embargo, si el conflicto explícito involucra observación y si lo que nos interesa son las inconsistencias entre teoría y observación, las características de los resultados observacionales no puede desestimarse sólo porque el criterio de independencia no ha sido plenamente satisfecho.

> **Contradicción T-O (Aux):** Dada una teoría empírica *T* que tiene a *s* como consecuencia observacional, si un experimento se realiza, se reporta ¬*s*; empero hay una importante intersección entre los supuestos relevantes de *T* y los supuestos relevantes de las teorías auxiliares involucradas en llevar a cabo la observación –incluyendo el diseño de instrumentos, la interpretación del resultado observacional y/o del diseño del experimento.

En economía experimental y del comportamiento se desarrollaron los juegos del dictador y del ultimatum que generaron un fuerte cuestionamiento a la idea de que los agentes racionales son individualistas y calculadores. El juego del dictador consiste en que se tiene dos jugadores (A y B). Al jugador A se le da una cantidad de dinero y se le dice que la debe repartir entre él y el jugador B. El jugador B no puede participar en la fijación del monto de la cantidad repartida. El jugador A dicta la cantidad y esa es la que se reparte. Cuando se trataba de jugadores anónimos, el jugador A repartía cantidades muy desiguales (la mayoría repartía una proporción entre 90% y 10% del total de recursos). Esto corroboraba la conducta de agente racional ortodoxo. Pero cuando se pasaba al juego del ultimatum, donde todo se introducía la posibilidad de que el jugador B pueda aceptar o rechazar la oferta del jugador A. Si la oferta era aceptada, el monto de dinero repartido era el que se había propuesto por A. Si la oferta era rechazada ninguno de los dos se quedaba con dinero.

Cuando se ofrecían cantidades desiguales (90 % y 10 %) el jugador B prefería rechazar la oferta. Esto hace que conducta de B fuera contraria a la predicción de la economía ortodoxa, este prefería castigar al jugador A por presentar propuestas inequitativas a con ello asumir que tendría un monto

igual a cero. Esto significaba que no maximizaba su utilidad monetaria, donde la conducta esperada de un homo economicus sería aceptar la oferta del 10% de dinero en vez de nada. En cambio, el jugador B tendía a aceptar las ofertas más igualitarias (50% - 50%). Esto sugería que los agentes, en muchos contextos, tomaban decisiones económicas motivados no sólo por aspectos monetarios, sino por elementos emocionales, como es el sentimiento de justicia y lo que en la literatura conductual se conoce como aversión a la inequidad (cf. KRUEGER, GRAFMAN Y MCCABE 2008).

Es importante notar que distinguir entre distintos subtipos de contradicciones entre teoría y observación no significa que las anomalías aquí descritas no sean contradicciones lógicas, sólo significa que algunas contradicciones lógicas que involucran conflictos entre teoría y observación no son tan simples como tendemos a imaginar, y que para estudiarlas y obtener la máxima cantidad de información sobre las teorías en cuestión y sobre la ciencia misma, debemos ser capaces de diferenciar los distintos tipos.

Para los economistas puede ser relevante conocer las herramientas de la lógica paraconsistente para evitar pensar que su racionalidad (o la de otros economistas) está en peligro cuando en un mismo modelo o teoría se asumen supuestos que pueden ser contradictorios entre si. Cuando se trata de contradicciones no explosivas, los economistas pueden tomar una actitud más tolerante hacia el uso de supuestos contradictorios entre si. Un ejemplo, como ya hemos dicho, es asumir que los agentes son perfectamente racionales y, en otras ocasiones, asumir que no son perfectamente racionales debido a su racionalidad acotada y sus sesgos cognitivos. En apariencia puede parecer contradictorio. Pero no necesariamente es así, porque el supuesto de racionalidad perfecta puede ser útil en ciertos momentos argumentales, como vimos con los ejemplos sobre corrupción e incentivos. En esos casos se pueden predecir comportamientos en contextos específicos. En otros momentos argumentales se puede asumir agentes con racionalidad acotada y sesgos cognitivos para incorporar al análisis elementos psicológicos que ayuden a analizar razonamientos morales poco éticos. El uso de los supuestos, por tanto, depende del objetivo que se busca alcanzar con ellos.

La tolerancia entre las contradicciones entre enunciados teóricos y enunciados observacionales también puede servir a los economistas para adoptar actitudes más cuidadosas sobre cómo interpretar la evidencia empírica que tienen a su disposición, así como sus consecuencias sobre una teoría. Es decir, puede haber situaciones donde una evidencia empírica de un contexto específico contradiga a una teoría o modelo específico. Pero eso no significa que el modelo siempre va a estar mal. De hecho puede haber muchos otros contextos donde el modelo funcione bien. Asumir una posición más tolerante hacia las contradicciones no debe llevarnos a sobrevalorar la evidencia empírica de un contexto. Las inferencias deben hacerse con más cuidado. Por ejemplo, que haya evidencia empírica contra el modelo estándar de elección racional no nos lleva a desechar ese modelo en todos los casos. Más bien, tolerar la contradicción en este caso, nos lleva a adoptar una actitud más humilde sobre los rangos de aplicación de ese y otros modelos económicos. Nos ayuda a saber de una forma más clara qué se puede y qué no se puede hacer con un modelo.

Finalmente, tolerar la contradicción entre teorías, y entre teoría y observación, puede llevar a los economistas a evitar proyectos poco fructíferos como insistir en unificar lo que, en principio parece no-unificable. Es el caso de intentar unificar economía neoclásica y economía del comportamiento dentro de un marco teórico más amplio. Tener este marco teórico que busque resolver contradicciones teóricas no necesariamente constituye un avance científico, pues no es claro que la unificación sea en sí misma algo valioso. De nada sirve tener una teoría que no tiene contradicciones internas, si no queda claro qué problemas resuelve. De hecho, la unificación por sí misma parece carecer de sentido. Es como tener una respuesta que no responde a ninguna pregunta. Al igual que en física, en economía se puede trabajar de manera fructífera tolerando contradicciones para realizar modelos con capacidad predictiva, o para elaborar diseños de política económica que sí resuelvan problemas específicos. Tolerar contradicciones puede ayudar a evitar comportamientos dogmáticos y fundamentalistas entre los economistas, para llevarlos a tener posiciones más abiertas.

6. Reflexiones finales

A lo largo de este texto caracterizamos la tolerancia a las contradicciones en la ciencia como es la práctica de poder razonar ya sea a partir de o con información inconsistente, y seguir siendo capaz de identificar las consecuencias de nuestro razonamiento que son sensatas en determinados contextos de las que no lo son. Defendimos que dicha tolerancia puede ocurrir en tres niveles distintos: a nivel doxástico, a nivel del razonamiento de los agentes, y a nivel de las teorías. Nos concentramos en los casos que ocurren a nivel del razonamiento y discutimos bajo qué condiciones su estudio puede arrojar luz con respecto a los patrones inferenciales que permiten la tolerancia a las contradicciones en las ciencias –sean éstos paraconsistentes o no. Para hacerlo, consideramos uno de los argumentos más usados en la literatura para motivar la aplicabilidad de las lógicas paraconsistentes en las ciencias y explicamos cuáles eran los vicios metodológicos de dicho argumento, siendo el más importante el (injustificado) supuesto de uniformidad de los casos de tolerancia a la inconsistencia en las ciencias. Posteriormente, propusimos un camino metodológico que consistía en cuestionar dicha uniformidad y en explorar las diferencias particulares de los distintos casos.

Defendimos que las contradicciones entre teoría y observación y las prácticas de tolerancia a la inconsistencia que tienen lugar en las ciencias sociales ofrecen un terreno fértil para el estudio de la tolerancia racional a las contradicciones. Ofrecimos dos estudios de caso que ilustran la tolerancia de dos contradicciones entre teoría y observación, uno de ellos, proveniente de la física y el otro de la economía. Después de evaluar ambos casos, discutimos las similitudes entre ambos y la importancia que tiene su estudio para la identificación de mecanismos, tanto inferenciales como epistémicos, para la tolerancia racional a las contradicciones.

Posteriormente, presentamos una tipología de las contradicciones entre teoría y observación que es posible encontrar en las ciencias. Esta tipología incluye tres subtipos de contradicciones:

- las que satisfacen un criterio de independencia observacional (contradicción T-O (Indp)),
- las que necesitan ser evaluadas desde un marco de trabajo adicional para ser reconocidas como contradicciones (Contradicción T-O+M), y
- aquellas no satisfacen el criterio de independencia observacional por el holismo que permea las relaciones entre la teoría que está siendo evaluada, la observación y la teoría que subyace al reporte observacional (Contradicción T-O (Aux)).

Dado que las contradicciones entre teoría y observación son tanto las más comunes como las mejor toleradas en las ciencias, su estudio detallado permitirá alcanzar una mejor comprensión de la tolerancia a la inconsistencia en las ciencias. En particular, el estudio de subtipos de contradicciones entre teoría y observación permitirá identificar distintas estrategias de razonamiento empleadas para preservar el razonamiento sensato en contextos inconsistentes.

Un estudio sustancial de los mecanismos de tolerancia a la inconsistencia en las ciencias requiere prestar atención no solamente a las distintas contradicciones sino también a la forma en la que éstas se comportan en diferentes disciplinas científicas. Por esta razón, parte importante de la contribución de este artículo ha sido el dirigir la atención a las contradicciones en las ciencias sociales, en particular, a la economía.

Agradecimentos

Estamos agradecidos con Moisés Macías Bustos por los intercambios fructíferos sobre la pertinencia filosófica de esta discusión. Agradecemos a Alejandro Vázquez del Mercado por su ayuda durante la traducción al español de este texto. Estamos en deuda con Ana Rosa Pérez-Ransanz y Ambrosio Velasco-Gómez por recordarnos siempre la importancia de producir investigación en español. Agradecemos a los dictaminadores anónimos por sus valiosos comentarios, y a los editores del volumen por su enorme paciencia y tolerancia a nuestras contradicciones.

La primera autora agradece a Moisés Macías-Bustos y a Jean Yves Beziau por su constante apoyo durante la elaboración de este texto. Esta colaboración fue financiada por el Programa Nacional de Pós-Doutorado PNPD/ CAPES y el Proyecto UNAM PAPPIT IN403719 "Intensionalidad hasta el final: un nuevo plan para la relevancia lógica". El segundo autor agradece al Seminario de filosofía de la economía de la UNAM por permitirle profundizar en la literatura de economía experimental y del comportamiento. Su colaboración fue financiada por el Programa de Becas Posdoctorales en la UNAM, Centro de Ciencias de la Complejidad (asesorado por Enrique Cáceres Nieto), y el proyecto UNAM PAPIIT AV100120 "Salud y enfermedad: tecnologías de la información y biomarcadores no-invasivos de alertas tempranas".

Referencias

ARIELY Dan (2008), Predictably irrational . The hidden forces that shape our decisions ED. Harper, Nueva York, Estados Unidos.

BATENS, D. (2002): "In Defence of a Programme for Handling Inconsistency", in Inconsistency in Science, Kluwer Academic Publishers, Netherlands: 129-150.

BATENS, D. (2017): "Pluralism in scientific problem solving: why inconsistency is no big deal", HUMANA.MENTE Journal for Philosophical Studies 32: 149-177.

BECKER, G. S. (1993): Nobel lecture: The economic way of looking at behavior. Journal of political economy, 101(3), 385-409.

BÖHM-BAWER, E. (1949): Karl Marx and the close of his system, Ludwig von Mises Institute

BOWLES, S., y POLANIA-REYES, S. (2012): Economic incentives and social preferences: substitutes or complements? Journal of Economic Literature, 50(2), 368-425.

BROWN, B. y G. PRIEST (2004): "Chunk And Permeate, A Paraconsistent Inference Strategy. Part I: The Infinitesimal Calculus", Journal of Philosophical Logic 33: 379-88.

BROWN, B. y G. PRIEST (2015): "Chunk and permeate II: Bohr's Hydrogen Atom", European Journal for Philosophy of Science 5 (1): 1-18.

BUENO, O. (1997): “Empirical Adequacy: A Partial Structures Approach”, Studies in History and Philosophy of Science 28: 585-610.

BUENO, O. (2006): “Why Inconsistency Is Not Hell: Making Room for Inconsistency in Science”, en Knowledge and Inquiry: Essays on the Pragmatism of Isaac Levi, Cambridge Studies in Probability, Induction and Decision Theory, Cambridge University Press: 70-86.

CARNIELLI, W. y M. E. CONIGLIO (2016): Paraconsistent Logic: Consistency, Contradiction and Negation, Logic, Epistemology, and the Unity of Science, Springer.

DA COSTA, N. C. y S. FRENCH (2003)Science and partial truth: A unitary approach to models and scientific reasoning. Oxford University Press

DUHEM, P. (1991): The Aim and Structure of Physical Theory, Princeton University Press.

ELSAMAHI, M. (2005): “Coherence between theories”, Canadian Journal of Philosophy 35(2): 331-352.

ELSTER J. (2012), El desinterés. Tratado crítico del hombre económico, Siglo XXI. México

FRIEDMAN, M. (1953): “The methodology of positive economics”. Essays in positive economics, 3(3), 145-178.

FRIEDMAN, M. (1968): “The Role of Monetary Policy”, American Economic Review, (1968) 58(1).

FRISCH, M. (2004): "Inconsistency in Classical Electrodynamics", Philosophy of Science 71: 525-49.

GÄHDE, U. (2002): “Holism, underdetermination, and the dynamics of empirical theories”, Synthese 130: 60-90.

GINE, J. (2008): “On the origin of the anomalous precession of Mercury’s perihelion”, en Chaos, Solitons and Fractals Journal 38: 1004-1010.

GORDON, R. J. (2011): "The History of the Phillips Curve: Consensus and Bifurcation". Economica. 78 (309): 10–50.

HARPER, W. (2007): “Newton’s methodology and Mercury’s perihelion before and after Einstein”, Philosophy of Science 74 (5): 932-942.

HEIDL, S. (2016). Philosophical Problems of Behavioural Economics (1st ed.). Routledge.

HEMPEL, C. (1965): Aspects of Scientific Explanation and Other Essays in the Philosophy of Science, New York: Free Press.

HEMPEL, C. G. (2000): Selected philosophical essays. R. Jeffrey (ed.), New York: Cambridge University Press.

HOOVER, K. D. (2008): "Phillips Curve". en David R. Henderson (ed.). Concise Encyclopedia of Economics (2nd ed.). Indianapolis: Library of Economics and Liberty.

KRUEGER F., GRAFMAN J. y K. MCCABE (2008): "Neural correlates of economic game playing" en Philosophical Transactions of the Royal Society, 363 3859-3874

KUHN, T. (1977): The Essential Tension: Selected Studies in Scientific Tradition and Change, University of Chicago Press.

KUIPERS, T.A.F. (1999): "Abduction Aiming at Empirical Progress or even a Truth Approximation, Leading to Challenge for Computational Modelling", Foundations of Science 4 (3): 307-323.

KUIPERS, T. A. F. (2000): From Instrumentalism to Constructive Realism. On some relations between Confirmation, Empirical Progress and Truth Approximation, Synthese Library (vol. 287).

LAKATOS, I. (1965): Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science, Cambridge University Press, Cambridge, 1st edition.

LAKATOS, I. (1970): "Falsification and the Methodology of Scientific Research Programmes", en I. Lakatos, I. y A. Musgrave (eds.), Criticism and the Growth of Knowledge; Cambridge University Press: 91-195.

LAKATOS, I. (1978): The Methodology of Scientific Research Programmes: Philosophical Chapters, Vol. 1. Cambridge University Press.

LAUDAN, L. (1977): Progress and its Problems: Toward a Theory of Scientific Growth. Ewing, NJ: University of California Press.

LAZEAR, E. (2000), "Economic Imperialism", en The Quarterly Journal of Economics, Vol. 115, No. 1. Oxford University Press.

MANDEL, E. (1991). Introduction to volume 1 of Capital. In Capital: A critique of Political Economy (Vol. 1). Penguin Classics Books.

MARTÍNEZ-ORDAZ, M. del R. (2017): "Holism, Inconsistency Toleration and Inconsistencies between Theory and Observation", HUMANA.MENTE Journal of Philosophical Studies, 32: 117-147.

MARTÍNEZ-ORDAZ, M. del R. (2020): "The ignorance behind inconsistency toleration", S.I. Knowing the unknown, Synthese.

MARX, K. (1992a). Capital: A critique of Political Economy (Vol. 1). Penguin Classics Books.

MARX, K. (1992b). Capital: A Critique of Political Economy (Vol. 2). Penguin Classics Books.

MARX, K. (1991). Capital: A Critique of Political Economy (Vol. 3). Penguin Classics Books.

MAZAR N., A. O y ARIELY D. (2008): "The dishonesty of honest people: A theory of self-concept maintenance" en Journal of Marketing Research, Vol. XLV American Marketing Association.

MEHEUS, J. (2002): "How to reason sensibly yet naturally from inconsistencies", en Meheus, J. (ed) Inconsistency in Science: 151-164.

MOORE, G. E. (1942): "A reply to my critics" en The Philosophy of G. E. Moore, P. A. Schilpp (ed.). Evanston, IL: Northwestern University.

MORRISON, M. (2015): Reconstructing reality: models, mathematics, and simulations, Oxford: Oxford Studies in Philosophy.

NOWAK, L. (1980): The structure of idealization: Towards a systematic interpretation of the Marxian idea of science (Vol. 139). Springer Science & Business Media.

O' DONOGHUE, T., y RABIN, M. (1999): Doing it now or later. American economic review,89(1), 103-124.

O'NEIL, W.M. (1969): Fact and Theory: An aspect of the philosophy of science. Sydney University Press.

PHELPS, E. S. (1967): "Phillips Curves, Expectations of Inflation and Optimal Employment over Time." Economica, n.s., 34, no. 3: 254–281.

POPPER, K. (1959): The Logic of Scientific Discovery. London: Hutchinsonn and Co.

PRIEST, G. (2002)Inconsistency and the empirical sciences. See Meheus (2002), pp. 119–128.

PRIEST, G. (2002): "Inconsistency and the Empirical Sciences", in Inconsistency en Science, Kluwer Academic Publishers, Netherlands: 119-28.

ROSE-ACKERMAN S. (1974): "The economics of corruption", Journal of Public Economics 4 North-Holland Publishing Company.

ROVANE, C. (2004): "Rationality and Persons", en Piers Rawling y Alfred R. Mele (eds.), The Oxford Handbook of Rationality. Oxford: Oxford University Press: 320-342.

ŠEŠELJA, D. (2017). Scientific Pluralism and Inconsistency Toleration. HUMANA.MENTE Journal of Philosophical Studies, 10(32), 1-29.

ŠEŠELJA, D. (2017): "Scientific Pluralism and Inconsistency Toleration", HUMANA.MENTE Journal for Philosophical Studies 32: 1-29.

SWEENEY, D. C. (2014): "Chunk and Permeate: The Infinitesimals of Isaac Newton", History and Philosophy of Logic, 35: 1-23.

THALER R. (1980), "Toward a positive theory of consumer choice", Journal of Economic Behavior and Organization, Vol. 1, North Holland.

THALER R. (2000): "From homo economicus to homo sapiens", en The Journal of Economic Perspectives, Vol. 14, Num. 1, American Economic Association.

TVERSKY A. y KAHNEMAN D. (1974): "Judgment under uncertainty: Heuristics and biases", Science, New Series, Vol. 185 (4157): 1124-1131.

VICKERS, P. (2013): Understanding Inconsistent Science, Oxford University Press.

ESTICANDO A VERDADE SEM ALCANÇAR A CONTRADIÇÃO: O CASO DO DIALETEÍSMO

Ederson Safra Melo[1]

Jonas Rafael Becker Arenhart[2]

RESUMO

Dialeteias são contradições verdadeiras. Dialeteísmo é a visão de que há dialeteias e dialeteístas são aqueles que defendem tal visão. Uma das principais motivações para o dialeteísmo encontra-se nos paradoxos semânticos, como o paradoxo do Mentiroso. A sentença do mentiroso, na abordagem dialeteísta, instancia uma dialeteia. O problema é: como aceitar contradições verdadeiras? Como dialeteísmo é a visão que *algumas,* mas não *todas,* contradições são verdadeiras, o dialeteísmo demanda um tratamento paraconsistente. Porém, não é qualquer sistema paraconsistente que pode ser aplicado ao dialeteísmo. A interpretação do símbolo de negação deve ser um operador formador de contradição (ofc) com o sentido relevante para a aplicação dialeteísta. Especificamente, na versão de dialeteísmo que discutiremos nesse artigo, defendida por Graham Priest, temos que, para que o dialeteísmo faça sentido, a negação da lógica subjacente deve atender a dois requisitos: (i) ser um ofc e (ii) não ser explosiva. A Lógica do Paradoxo (LP) tem sido apontada como a lógica adequada ao dialeteísmo. Veremos que, a fim de lidar com contradições sem trivialidade, dialeteístas "esticam" a verdade de modo que a verdade possa incluir, em alguns casos, a falsidade também. Nesse caso, haveria sentenças verdadeiras e falsas, chamadas de *aglutinações de valores de verdade* (*truth-value gluts*). Em LP, aglutinações de valores de verdade são fundamentais para garantir os requisitos (i) e (ii). Todavia, vamos argumentar que tal procedimento de esticar a verdade, permitindo aglutinações, garante a paraconsistência ao custo de distorcer a interpretação da noção de contradição envolvida no dialeteísmo. Especificamente, tal distorção enfraquece a interpretação da negação comprometendo o sentido de contradição relevante para o dialeteísmo. Vamos argumentar que as próprias restrições dialeteístas a uma compreensão adequada do Mentiroso mostram que as condições (i) e (ii) são incompatíveis e que, com isso, o projeto dialeteísta enfrenta consideráveis obstáculos.

Palavras-chave: Dialeteísmo. Contradição. Paraconsistência. Negação. Paradoxo do Mentiroso.

[1] Professor do Departamento de Filosofia da Universidade Federal do Maranhão, Brasil. E-mail: edersonsafra@yahoo.com.br. ORCID: 0000-0003-1706-1128.

[2] Professor do Departamento de Filosofia da Universidade Federal de Santa Catarina, Brasil. Bolsista de Produtividade em Pesquisa do CNPq. Artigo produzido enquanto o autor era *Research Fellow* no Instituto Círculo de Viena, na Universidade de Viena, Áustria. E-mail: jonas.becker2@gmail.com. ORCID: 0000-0001-8570-7336.

ABSTRACT

Dialetheias are true contradictions. Dialetheism is the view that there are true contradictions, and dialetheists are those defending such view. One of the main motivations for dialetheism is found in the semantic paradoxes, among them, the Liar paradox. The Liar sentence, in the dialetheist approach, instantiates a dialetheia. The problem that remains to be solved is: how to accept true contradictions? Given that dialetheism is the view that *some,* but not *all* contradictions are true, a paraconsistent logic will be required. However, it is not the case that any paraconsistent logic will work. The interpretation of the negation symbol must be a contradictory forming operator (cfo) with the relevant meaning for the dialetheist case. Specifically, according to the version of dialetheism we shall focus on in this paper, Graham Priest's dialetheism, in order for dialetheism to make sense, the underlying logic's negation must satisfy two conditions: (i) be a cfo and (ii) avoid explosion. The Logic of Paradox (LP) has been frequently pointed as the appropriate system for dialetheism. As we shall see, in order to deal with contradictions without triviality, dialetheists "stretch" the truth so that truth includes, in some cases, falsity too. In this case, there could be sentences that are true and false, *called truth-value gluts.* In LP, gluts are essential to grant the success of conditions (i) and (ii). However, we shall argue that such a procedure of stretching the truth, granting gluts, delivers paraconsistency at the price of distorting the required meaning of contradiction involved in dialetheism. Specifically, such a distortion weakens the notion of negation, threatening the meaning of contradiction that is essential for dialetheism. We shall argue that the restrictions imposed by dialetheists to an appropriate understanding of the Liar show conditions (i) and (ii) are incompatible, and as a result, the dialetheist project faces considerable obstacles.

Keywords: Dialetheism, Contradiction, Paraconsistency, Negation, Liar paradox.

1. Introdução

Contradições podem aparecer em diversas situações, que vão do senso comum aos contextos científicos. Diante desse fato, a pergunta que surge naturalmente é: qual deve ser nossa atitude diante de tais contradições? Podemos aceitar algumas delas? A resposta mais influente na tradição filosófica para esta questão é negativa, e foi dada por Aristóteles na Metafísica, ao defender a *Lei da Não Contradição* (LNC). Segundo Aristóteles, LNC conta como sendo o mais certo de todos os princípios. Assim, de acordo com uma formulação *semântica* do LNC, uma sentença, α, e sua negação, $\neg\alpha$, não po-

dem ser ambas verdadeiras (existem também formulações *psicológicas* e *ontológicas* do princípio; aqui, focaremos na formulação semântica). Apesar de LNC encontrar alguma oposição na história da filosofia durante os séculos, permanece ainda hoje como parte da visão hegemônica. Mais recentemente, já na segunda metade do século XX, um novo desafio foi lançado por uma abordagem batizada de *dialeteísmo,* que defende que algumas contradições são verdadeiras, opondo-se, nesse sentido, à LNC. Graham Priest, um dos principais defensores do dialeteísmo, chega ao ponto de caracterizar dialeteísmo como uma oposição explícita a LNC: "[a] visão que LNC falha, que algumas contradições são verdadeiras, é chamada dialeteísmo" (PRIEST, 2004, p. 29).

A defesa de Aristóteles de que contradições não podem ser verdadeiras parece ter exercido uma grande influência no desenvolvimento do pensamento ocidental, incluindo a lógica e a filosofia. Essa defesa contribuiu, segundo Priest, para que LNC fosse tomada como certa, evitando uma discussão subsequente sobre a aceitabilidade de LNC. De acordo com os dialeteístas, o fato de poucos filósofos na história da filosofia terem questionado a defesa aristotélica de LNC não significa que os argumentos de Aristóteles sejam conclusivos. Para Priest, tais argumentos não funcionam (cf. PRIEST, 2002, p. 5), e se baseiam mais na autoridade de Aristóteles do que em seus próprios méritos. Sobre esse ponto, Priest diz o seguinte:

> Eu suspeito (infelizmente) que a visão foi aceita simplesmente com base na autoridade magistral dos textos de Aristóteles na Idade Média. Em geral, essa autoridade desapareceu há muito tempo, é claro. Na lógica, durou até o século XX; a maior parte foi varrida desde então, mas as opiniões sobre a contradição se mantiveram obstinadamente. (PRIEST, 2006b, p. 121).

Mas ainda que a defesa aristotélica de LNC seja falha — o que é um assunto que gera um grande debate, mas que não abordaremos aqui –– em um contexto lógico, parece haver um excelente motivo para que contradições sejam evitadas: seguindo princípios de inferência bastante simples, a partir de uma contradição seríamos levados a concluir que todas as sentenças são verdadeiras.

1. Assuma que $\alpha \wedge \neg \alpha$ é verdadeira;
2. de (1) e simplificação, temos α;
3. de (2) e expansão, temos $\alpha \vee \beta$;
4. de (1) e simplificação, temos $\neg \alpha$;
5. de (2), (3) e silogismo disjuntivo, temos β.

Essa derivação é válida em várias lógicas como, por exemplo, na lógica clássica e na lógica intuicionista. Um sistema de lógica na qual podemos derivar que qualquer sentença é verdadeira é dito possuir uma relação de consequência *explosiva* ou *trivial*. Especificamente, vale o princípio da explosão, ou *ex falso quodlibet*, que intuitivamente diz que a partir de uma contradição, $\alpha \wedge \neg \alpha$, podemos inferir qualquer sentença, β.

Tendo isso em vista, aceitar uma contradição implicaria em aceitar o trivialismo (a tese de que todas as sentenças são verdadeiras). Um dialeteísta, ao aceitar contradições verdadeiras, não seria também um trivialista? É fundamental que o dialeteísta evite uma relação de consequência lógica explosiva caso se queira coerente, e esse é um dos aspectos fundamentais do dialeteísmo. Segundo os dialeteístas, a inferência que leva de contradições para a explosão comete uma petição de princípio e, portanto, não poderia funcionar contra a posição que defende que há contradições verdadeiras (cf. PRIEST, 2019a, p. 681). Como o dialeteísmo é a visão de que algumas, mas *não todas*, contradições são verdadeiras, a explosão deve falhar no contexto dialeteísta. Lógicas que violam o princípio de explosão, por definição, são paraconsistentes. Ou, como é frequentemente dito, lógicas paraconsistentes são "lógicas em que contradições não implicam qualquer coisa" (PRIEST, 2007a, p. 129).

Um dos principais fatores para o desenvolvimento do dialeteísmo foi o advento dos sistemas paraconsistentes a partir da metade do século XX. Priest (2019a, p. 682) afirma que o dialeteísmo não teria sido levado a sério sem o desenvolvimento das lógicas paraconsistentes. Existe hoje uma grande variedade de sistemas paraconsistentes, todos eles invalidando a lei de explosão. Diante dessa multiplicidade, os dialeteístas colocam restrições acerca de qual

sistema paraconsistente é aceitável; ter uma lógica em que $\alpha \wedge \neg \alpha$ não leva à trivialidade não é suficiente para o dialeteísmo. A lógica paraconsistente, no âmbito dialeteísta, deve formar contradições; o operador de negação envolvido deve ser um operador formador de contradição (ofc, daqui em diante). O sentido de contradição envolvida é relevante! Consideramos que a aplicação da paraconsistência como suporte ao dialeteísmo deixa claro qual o sentido da contradição que deve ser capturado pela lógica subjacente à teoria que pretende lidar com contradições verdadeiras.

O dialeteísmo tem se destacado como uma teoria capaz de lidar com paradoxos semânticos, como o paradoxo do Mentiroso. No contexto dialeteísta, além de evidenciar que há contradições verdadeiras (instanciando, assim, a definição de dialeteísmo), o paradoxo do Mentiroso explicita o sentido da contradição requerido pelos dialeteístas. Desse modo, o paradoxo pode ser tomado como uma espécie de guia para a noção de contradição relevante para o dialeteísmo. Levando isso em conta, para que a posição dialeteísta faça sentido, a lógica subjacente ao dialeteísmo deve cumprir conjuntamente duas exigências: (i) ter uma negação formadora de contradição (com o sentido relevante para a aplicação dialeteísta) e (ii) não permitir a explosão a partir de uma contradição. A *Lógica do Paradoxo* (LP) tem sido tomada como o sistema adequado para manter contradições verdadeiras sem trivialidade. Todavia, apesar de ser à primeira vista compatível com todas as exigências dialeteístas, vamos argumentar que a lógica subjacente ao dialeteísmo não pode manter ambas exigências (i) e (ii) e que, portanto, o dialeteísmo enfrenta desafios em sua própria articulação. Tendo isso em vista, nosso procedimento, nesse artigo, será o seguinte: faremos primeiramente uma apresentação do dialeteísmo, e depois apontaremos alguns problemas para a tese que afirma que há contradições verdadeiras sem trivialidade. A interpretação da contradição em LP, tendo em vista a sua aplicabilidade no dialeteísmo, estará no foco deste artigo. Voltamos a enfatizar que embora muitos argumentos desenvolvidos no presente artigo possam ser vistos como obstáculos para outras formas de dialeteísmo, vamos restringir nossas considerações para a forma mais conhecida e desenvolvida de dialeteísmo, o dialeteísmo de Priest.

A estrutura do presente artigo é dividida da seguinte maneira. Na seção 2, vamos expor os principais conceitos do dialeteísmo e a sua relação com a lógica paraconsistente. Especificamente, trataremos da relação entre dialeteísmo e paraconsistência, conferindo os requisitos necessários para que uma teoria paraconsistente possa ser adequada, segundo os critérios dialeteístas. Faremos uma apresentação breve da Lógica do Paradoxo (LP), que usualmente é defendida como um sistema adequado para lidar com contradições verdadeiras sem trivialidade. Na seção 3, vamos tratar de uma das principais motivações para o dialeteísmo: o *paradoxo do Mentiroso*. Particularmente, vamos conferir a característica do Mentiroso, chamada por Priest de "essência do Mentiroso", que sobrevive às várias formulações do paradoxo através de suas formas estendidas. Na linha da argumentação dialeteísta, veremos que "a essência do Mentiroso" revela o sentido pré-teórico da contradição que deve ser capturado pela negação de LP, enquanto um operador formador de contradição. Na seção 4, vamos estabelecer a relação entre a essência do Mentiroso e o comportamento que se espera que a negação tenha em LP. A partir daí, através de uma conhecida crítica de Slater à paraconsistência, vamos argumentar que o comportamento da negação paraconsistente retrata uma noção mais fraca do que aquela incorporada na essência do Mentiroso (que interpreta a negação como um ofc). Vamos apresentar os argumentos de Slater e focar na resposta de Priest a favor da ideia de que a negação de LP é interpretada como um ofc e que LP garante a essência do Mentiroso (i.e., o sentido de contradição adequado ao dialeteísmo). Na seção 5, vamos dar um passo além no debate Priest-Slater e defender que a resposta de Priest aos argumentos de Slater não pode ser sustentada. Especificamente, argumentaremos que a negação de LP não apenas não é um ofc, mas ainda mais do que isso: a negação de LP não pode ser interpretada como um ofc, tendo em vista a aplicação pretendida pelo dialeteísta. Vamos argumentar que a negação de LP só será não-explosiva ao custo de não ser um ofc (com a interpretação da contradição presente no Mentiroso). A partir daí apontaremos alguns resultados que questionam o sucesso do dialeteísmo enquanto a tese de que há contradições verdadeiras. Concluímos na seção 6.

2. Dialeteísmo e paraconsistência: conceitos básicos.

O termo 'dialeteísmo' foi cunhado por Graham Priest e Richard Routley (Sylvan posteriormente) em 1981. A inspiração para o nome reside nas observações de Wittgenstein, sobre os Fundamentos da Matemática, onde ele descreve o paradoxo do Mentiroso como uma figura de dupla face (*Janus-headed*) diante da verdade e da falsidade. Com base nessa inspiração, uma dialeteia seria uma espécie de verdade de duas vias. Entretanto, Routley e Priest tomaram caminhos diferentes acerca de como melhor grafar o neologismo. Routley começou a escrever sem o 'e' ('*dialethism*'), enquanto Priest começou a escrever com o 'e' ('*dialetheism*'). Talvez por conta do livro *In Contradiction* — publicado pela primeira vez em 1987 e conhecido posteriormente como um "manifesto canônico do dialeteísmo" —[3] a grafia 'dialeteísmo' ('*dialetheism*') tenha se tornado a maneira padrão para chamar a doutrina que sustenta que há contradições verdadeiras (cf. PRIEST, 2019a, p. 683).

Embora a teoria dialeteísta tenha sido batizada relativamente há pouco tempo (para os padrões de tempo da filosofia), dialeteístas defendem que se pode encontrar teses dialeteístas espalhadas pela história da filosofia desde os filósofos pré-socráticos até os dias atuais. Essas teses estariam espalhadas tanto em propostas explícitas de violação da LNC, quanto em contradições internas de sistemas dos grandes filósofos da tradição ocidental, que requerem uma violação de LNC para que se articule suas propostas. Em particular, para eles, há relatos dialeteístas explícitos na história da filosofia ocidental, como em Heráclito e em Hegel, e na filosofia oriental, com base em aspectos do Budismo.[4] Todavia, parece razoável afirmar que o dialeteísmo só se consolidou como uma tese lógico-semântica com os trabalhos de Priest, tendo sido desenvolvido também por outros autores, como Jc Beall. Conforme já mencionamos, vamos tratar aqui do dialeteísmo contemporâneo, focado no dialeteísmo de Priest.[5]

---

[3] Cf. PRIEST, 2006a p. xx.

[4] Para a concepções de Hegel e do Budismo, no contexto dialeteísta, ver, respectivamente, PRIEST, 2002, cap. 7 e PRIEST, 2002, cap. 16.

[5] Para aspectos históricos, ver PRIEST, G. ROUTLEY, R. 1989; PRIEST, 2007a, sec. 3; PRIEST, BERTO, WEBER (2018). Não vamos nos ocupar com questões históricas do dialeteísmo neste artigo. Nosso interesse aqui reside na abordagem dialeteísta

Anteriormente, seguindo a definição padrão, apresentamos o dialeteísmo como a visão de que há dialeteias. Uma dialeteia é uma contradição verdadeira; i.e., uma sentença (ou, mais genericamente, um portador de verdade), α, tal que tanto ela quanto sua negação, $\neg\alpha$, são verdadeiras. Como a falsidade de uma sentença é entendida como verdade da negação dessa mesma sentença, algumas sentenças (as dialeteias) resultam verdadeiras e falsas. Assim, há aglutinações (*gluts*) de valores de verdade. Na visão dialeteísta temos, portanto, que os conceitos de contradição, dialeteias e aglutinação (*gluts*) estão intimamente relacionados. A íntima relação de tais conceitos centrais aparece por quase toda a literatura dialeteísta. Vejamos rapidamente algumas passagens com a caracterização desses conceitos:

> Dialeteísmo é a visão de que algumas contradições são verdadeiras: há sentenças (enunciados, proposições ou qualquer coisa tomada como um portador de verdade), α, tal que ambas α e $\neg\alpha$ são verdadeiras, isto é, tal que α é tanto verdadeira quanto falsa (PRIEST, 2006b, p.1).

> Uma dialeteia é um par de sentenças (ou a conjunção delas, se preferir) da forma α e $\neg\alpha$, tal que ambas são verdadeiras. Se tomamos 'falso' sendo uma negação verdadeira, isso significa que α é tanto verdadeira quanto falsa. Dialeteísmo é a tese de que algumas αs são dialeteias (PRIEST, 2009b, p. 558).

> O que os Mentirosos nos ensinam, na visão dialeteísta, é que a verdade é inconsistente – que algumas sentenças verdadeiras têm negações verdadeiras. (BEALL, 2007, p. 05)

Para ficar mais claro, vamos organizar as noções extraídas das citações acima da seguinte maneira:

- **Contradições:** sentenças da forma 'α e $\neg\alpha$' (ou '$\alpha \wedge \neg\alpha$').[6]
- **Dialeteia**: uma contradição verdadeira, isto é, uma sentença α, tal que ambas α e $\neg\alpha$ são verdadeiras,
- **Aglutinação** (***glut***) d**e valor-verdade**: um valor de verdade que pode ser atribuído a uma sentença, significando que ela é verdadeira e falsa.

contemporânea.

[6] Veremos abaixo que o comportamento da negação terá um papel crucial para o dialeteísmo.

As noções de contradição e contradição verdadeira (dialeteia) são distintas e não equivalentes mesmo para um dialeteísta; no dialeteísmo, segundo a versão de Priest, poucas contradições são consideradas dialeteias, ou, em outras palavras, *as dialeteias são raras*. Como nas citações acima, algumas sentenças, como a sentença do Mentiroso, são casos de dialeteias e, considerando a noção de falsidade, elas são aglutinações de valores de verdade.[7] Assim, embora distintas, as noções de dialeteia e aglutinação resultam equivalentes. Como veremos, essa equivalência será importante para o tratamento paraconsistente dado às dialeteias.

Para manter que algumas contradições são verdadeiras sem cair no trivialismo, o dialeteísmo leva à paraconsistência. Todavia, a paraconsistência não leva ao dialeteísmo, na medida em que assumir uma teoria paraconsistente não implica em assumir que há contradições verdadeiras. Assim, embora relacionadas no contexto dialeteísta, é importante notar que dialeteísmo e paraconsistência são noções distintas. O dialeteísmo é uma tese bastante ampla sobre o comportamento da verdade e da negação (que sustenta a verdade de algumas contradições), enquanto que a paraconsistência diz respeito a noção de consequência lógica (que impede que qualquer sentença β seja consequência de $\alpha \wedge \neg \alpha$). Várias abordagens paraconsistentes aceitam sentenças da forma $\alpha \wedge \neg \alpha$ sem assumir que elas sejam verdadeiras, ou sem assumir que sejam aglutinações de valores de verdade. Por exemplo, os sistemas C_n de da Costa e as Lógicas da Inconsistência Formal (LFIs) não se comprometem com contradições verdadeiras e nem com aglutinações de valores de verdade (cf. CARNIELLI & CONIGLIO, 2016).

Para se compreender essa distinção, vale a pena, mesmo que brevemente, introduzir aqui uma discussão acerca da ideia de que há algo como uma interpretação de um determinado sistema paraconsistente em questão, levando em conta o seu campo de aplicação. Além do dialeteísmo, temos outras aplicações para a paraconsistência. Por exemplo, Carnielli e Rodrigues (2019) defendem uma abordagem epistêmica da paraconsistência, sustentando que devemos interpretar sentenças da forma $\alpha \wedge \neg \alpha$ em tais lógicas a partir da no-

[7] Vamos explorar o paradoxo do Mentiroso, no contexto da teoria dialeteísta, na próxima seção.

ção de evidência (uma noção mais fraca que a verdade). De acordo com essa abordagem, podem existir situações em que há evidências tanto para α quanto para $\neg\alpha$ (por exemplo, pode haver uma situação em que uma pessoa, ou um banco de dados, receba uma informação, α, de uma fonte e, também, receba uma informação, $\neg\alpha$, de outra fonte). Uma sentença $\alpha \wedge \neg\alpha$ pode ser aceita, com base em evidências, sem ser considerada verdadeira (contradições verdadeiras são banidas em tal abordagem, a verdade é consistente segundo Carnielli e Rodrigues). Para lidar com essa noção de maneira formal, Carnielli e Rodrigues apresentam dois sistemas paraconsistentes: BLE (*Basic Logic of Evidence*) e LET_j (*Logic of Evidence and Truth*). BLE é uma formulação equivalente da Lógica de Nelson (N4), aplicada ao tratamento lógico da evidência, e LET_j consiste em uma extensão de BLE com um operador de classicalidade resultando em uma lógica LFI. Com tal formalismo, quando temos um conflito de evidências (i.e., evidências para α e para $\neg\alpha$), temos um contexto paraconsistente, evitando a trivialidade. Por outro lado, quando é a noção de verdade que está em jogo, o contexto é clássico, e $\alpha \wedge \neg\alpha$ leva à trivialidade, uma vez que vale a explosão. Portanto, considerando a interpretação pretendida de evidências, não há contradições verdadeiras na abordagem epistêmica da paraconsistência (cf. CARNIELLI & RODRIGUES, 2017).

Tendo isso em vista, a depender da interpretação dada para os conectivos e, por consequência, para expressões sintáticas como $\alpha \wedge \neg\alpha$, uma questão que se levanta diz respeito ao significado de tal expressão, e sua consequente capacidade de representar uma contradição. Como a paraconsistência tradicionalmente é entendida como envolvendo sistemas de lógica que permitem contradições sem trivialidade, a expressão '$\alpha \wedge \neg\alpha$' em uma lógica paraconsistente a princípio deveria representar uma contradição. Todavia, considerando que há aplicações distintas da paraconsistência (dialeteísta e epistêmica, por exemplo), acompanhada de distintas interpretações para a negação, uma questão interessante é se alguma delas representa uma contradição no sentido pretendido, de violar LNC.

Para considerar essa questão mais adequadamente, aparentemente teríamos que ter uma noção prévia de contradição — uma compreensão infor-

mal da contradição. Ainda que a questão sobre aplicações distintas da paraconsistência e suas relações com as interpretações da contradição seja interessante, nesse artigo não faremos uma discussão específica sobre essa questão. Aqui vamos nos restringir ao dialeteísmo e, na seção 4, dentre outras coisas, trataremos da interpretação dialeteísta da noção de contradição. No entanto, a partir dessas considerações, fica claro também que sistemas paraconsistentes não se comprometem necessariamente com o dialeteísmo. Priest criou uma lógica paraconsistente, conhecida como Lógica do Paradoxo (LP), para lidar com dialeteias (a aplicação canônica de LP é o dialeteísmo). Mas nem mesmo LP está comprometida com o dialeteísmo. O próprio Priest (2019b, p. 558) menciona aplicações não dialeteístas de LP que dispensam completamente o valor "contraditório" como um valor de verdade. Pode-se oferecer, por exemplo, uma interpretação não dialeteísta para LP usando lacunas (*gaps*) de valores-verdade, o que é completamente banido pela interpretação canônica de LP.[8] No geral, sistemas paraconsistentes específicos não estão comprometidos com o dialeteísmo.[9]

Ainda que a paraconsistência de modo geral não esteja comprometida com a abordagem dialeteísta, tal abordagem precisa da paraconsistência, como apontamos acima. O dialeteísmo impõe condições bastante específicas sobre a formulação de uma lógica paraconsistente aceitável para seus objetivos. Além da paraconsistência (i.e. a violação da lei de explosão), o símbolo de negação, '¬', da teoria deve ser um operador formador de contradição (ofc). Em outras palavras, como dialeteísmo é a visão de que algumas contradições são verdadeiras, além de $\alpha \wedge \neg \alpha$ não levar ao trivialismo, essa fórmula deve ser interpretada como uma contradição, no sentido pretendido pré-teórico, conforme mencionamos acima. Ou seja, o símbolo de negação, '¬', da teoria deve ser um ofc. Mesmo com o risco de estarmos sendo repetitivos, é sempre importante enfatizar isso: para o dialeteísta, o significado da contradição captura-

[8] Intuitivamente, atribui-se a uma sentença uma lacuna de valores de verdade quando ela não recebe nenhum dos valores de verdade disponíveis, i.e., não é nem verdadeira e nem falsa. Como veremos, na interpretação dialeteísta de LP sugerida por Priest, não há lacunas de valores de verdade.

[9] Para uma discussão detalhada sobre as interpretações filosóficos da paraconsistência, veja ARENHART, 2020.

do pela lógica é fundamental![10] Segundo Priest, a lógica paraconsistente LP se mostra adequada aos propósitos dialeteístas. Vejamos, então, rapidamente, como LP funciona como uma lógica que pretende lidar com contradições verdadeiras sem trivialidade.

LP pode ser vista como uma lógica trivalente em que {*t, f, g*} é tomado como o conjunto de valores de verdade, onde '*t*' é interpretado como "verdadeiro e somente verdadeiro", '*f*' é interpretado como "falso e somente falso" e '*g*' é interpretado como "verdadeiro e falso". Nesse sentido, o valor *g* é entendido como uma aglutinação de valor-verdade (*truth-value glut*).[11] A aglutinação representa o caráter não exclusivo de LP; verdade e falsidade podem se sobrepor. Priest (1979) defende que uma das vantagens de LP, frente a outras lógicas paraconsistentes, reside no fato dela ter uma semântica bastante intuitiva.

> Tanto Routley (1977) quanto da Costa (1974) argumentaram que tais lógicas nos permitem investigar teorias inconsistentes, mas não triviais. Entretanto, nenhum desses sistemas propostos têm a simples e intuitivamente plausível semântica de LP (PRIEST, 1979, p. 226).

Em LP o comportamento dos operadores tem o mesmo comportamento da lógica clássica, quando restrito aos valores clássicos (ou seja, com os valores *t* e *f*): quando a entrada é clássica a saída é clássica. Mas como ficariam os casos que envolvem o valor *g*? Vamos ilustrar a situação intuitivamente, com apenas alguns casos particulares. Considere uma sentença α qualquer, para a qual foi atribuído o valor *g*. Qual seria o respectivo valor de $\neg\alpha$? Como α recebe o valor verdadeiro, $\neg\alpha$ terá o valor falso. Mas como α também tem o valor falso, $\neg\alpha$ terá o valor verdadeiro. Assim, aplicar a negação de LP em uma sentença, α, com o valor *g* resultará em uma sentença, $\neg\alpha$, também com o valor *g*. Agora vamos considerar o caso da conjunção $(\alpha \wedge \beta)$. Suponha que α tenha o valor *g* e β tenha o valor *t*. Como tanto α quanto β são verdadeiras, a conjunção será verdadeira. Mas α também é falsa

[10] Abaixo vamos voltar à discussão sobre o entendimento da noção de contradição no dialeteísmo.

[11] Na apresentação original de LP, em vez de 'g', Priest usa 'p', que ele chama de 'paradoxal' (cf. PRIEST, 1979).

e, assim sendo, a conjunção será falsa. Dessa forma, a conjunção $(\alpha \wedge \beta)$ terá o valor *g*. O raciocínio com a disjunção $(\alpha \vee \beta)$ é semelhante. Suponha que α tenha o valor *g* e β tenha o valor *f*. Como α é verdadeira, isso basta para que a disjunção $(\alpha \vee \beta)$ seja verdadeira. Mas ocorre que α é falsa, e β também. Assim, $(\alpha \vee \beta)$ resulta falsa. Dessa forma, a disjunção $(\alpha \vee \beta)$ terá o valor *g*. Tendo em vista o comportamento dos operadores, LP terá as seguintes tabelas:

	¬
t	*f*
g	*g*
f	*t*

∧	***t***	***g***	***f***
t	*t*	*g*	*f*
g	*g*	*g*	*f*
f	*f*	*f*	*f*

∨	***t***	***g***	***f***
t	*t*	*t*	*t*
g	*t*	*g*	*g*
f	*t*	*g*	*f*

As tabelas acima são as mesmas da lógica forte de Kleene (K_3). Assim como em K_3, em LP uma *valoração* é uma função que atribui um dos valores {*t, f, g*} para cada uma das fórmulas atômicas, e é estendida para todas as fórmulas da linguagem através do uso das cláusulas para as condições de verdade dos conectivos representados nas tabelas acima. Todavia, enquanto que em LP, o terceiro valor, denotado aqui por *g,* é interpretado como uma aglutinação de valor-verdade (*truth-value gluts*), em K_3 *g* é tradicionalmente interpretado como significando 'nem verdadeiro nem falso'; isto é, em K_3, *g* representa uma lacuna de valor-verdade (*truth value-gaps*). Ao assumir lacunas (*gaps*), em K_3, verdade e falsidade resultam como sendo não exaustivas. Além da interpretação do terceiro valor, outra diferença entre K_3 e LP está no conjunto de valores designados.[12] Assim como na lógica clássica, K_3 tem somente verdadeiro, *t*, como valor designado. Já em LP, tanto *t* quanto *g* contam como valores designados. Intuitivamente, em LP a verdade é "esticada" ou "alongada" para incluir os casos em que a verdade e a falsidade se sobrepõem, e essa verdade (que inclui algumas verdades falsas) é designada. Desse modo, a intuição subjacente para termos {*t, g*} como o conjunto de valores designados é que uma determinada sentença toma um valor designado quando ela é ao menos

[12] Lembrando que os valores designados são aqueles valores que são preservados em inferências válidas, e são os únicos valores que ocorrem em fórmulas válidas.

verdadeira (podendo eventualmente ser falsa também, mas nunca somente falsa). Dessa forma, ao "esticar" a verdade, incluindo *g* também como um valor designado, as noções usuais de validade e consequência lógica podem ser retratadas da seguinte maneira: uma sentença, α, será válida em LP ($\vDash \alpha$) quando for verdadeira em toda valoração (e isso inclui *g* que é verdadeira, ainda que também seja falsa). Similarmente, uma sentença, α, será consequência lógica semântica de um conjunto Γ de sentenças ($\Gamma \vDash \alpha$), se ela for verdadeira em toda valoração que torna os membros de Γ verdadeiros (incluindo *g* que é verdadeira, ainda que também seja falsa).

Tomando a noção usual de validade, e considerando a interpretação com lacunas de K_3, temos que K_3 não tem sentenças válidas. Particularmente, a formulação sintática do princípio do terceiro excluído (PTE), $\alpha \vee \neg \alpha$, não será válida em K_3.[13] Diferentemente, em LP, as formulações de PTE, $\alpha \vee \neg \alpha$ e LNC, $\neg(\alpha \wedge \neg \alpha)$, resultam válidas.[14] Considerando que LP e K_3 têm as mesmas tabelas, fica claro que o que faz LP validar certas fórmulas que não são válidas em K_3 é justamente a flexibilização do conjunto de valores designados, possibilitada pelo esticamento ou alongamento da verdade.

Parece um tanto curioso o fato da formulação sintática de LNC resultar válida em LP, apesar de o dialeteísmo ser precisamente uma tese questionando a validade LNC. Diante disso Priest diz que a lógica clássica é correta nos domínios consistentes, mas incorreta ao permitir que um "desastre" ocorra na presença de contradições.

> (...) uma importante e mesmo central parte da visão recebida da lógica deve ser reconhecida como errada: crucialmente, a lei de não contradição. Devemos ser cuidadosos aqui, entretanto, sobre o que exatamente isso significa. O que isso significa não é que a fórmula $\neg(\alpha \wedge \neg \alpha)$ não é logicamente válida (i.e., verdadeira em todas interpretações). De fato, essa fórmula é válida na semântica de LP. O que isso significa é que devemos aceitar algumas fórmulas $\alpha \wedge \neg \alpha$, uma vez que algumas delas são de fato verdadeiras (PRIEST, 1984, p. 154).

[13] Para conferir isso, tome o contra modelo em que α tem o valor *g* (lembrando que em K_3, *g* é interpretada como uma lacuna (*gap*) que, por sua vez, não é designado).

[14] É fácil conferir que $\alpha \vee \neg \alpha$ *e* $\neg(\alpha \wedge \neg \alpha)$ resultam válidas em LP, construindo tabelas para elas, considerando {*t, g*} como o conjunto de valores designados em LP. Voltaremos a considerar esse ponto mais abaixo.

Segundo os dialeteístas, o problema da lógica clássica está na validação da explosão. LP invalida a explosão através do uso do recurso de esticar a verdade. Com as aglutinações de valores de verdade sendo valores designados, é fácil construir um contra modelo que invalide a formulação mais tradicional da explosão $\alpha \wedge \neg \alpha \vDash \beta$. Considere o seguinte: que α receba um valor aglutinado (verdadeiro e falso) e β tenha o valor falso (e somente falso). Como α tem o valor *g*, pela tabela para a negação de LP, $\neg \alpha$ terá o valor *g* também. Dessa forma, a conjunção $(\alpha \wedge \neg \alpha)$ terá o valor *g*. Ora, como β tem o valor *f*, temos que $(\alpha \wedge \neg \alpha)$ recebe um valor designado e β recebe um valor não-designado, invalidando a explosão.[15] Portanto, temos que $\alpha \wedge \neg \alpha \nvDash_{LP} \beta$, garantindo assim o caráter paraconsistente de LP. Note que a paraconsistência só foi conseguida graças ao fato de que a verdade foi esticada, obtida através da permissão da aglutinação de valores de verdade. Assim, a noção de invalidade pode ser mantida: $\Gamma \nvDash \alpha$, quando os membros de Γ são (ao menos) verdadeiros e α falsa (e apenas falsa). Com isso, segundo os dialeteístas, a partir de LP, há recursos formais para representar contradições verdadeiras sem trivialidade.[16]

3. Contradições verdadeiras: o caso do Mentiroso

Como vimos, ainda que a paraconsistência seja necessária, ela não é suficiente para o dialeteísmo — dialeteístas precisam ainda nos dar razões para sustentar que há contradições verdadeiras. Em certo sentido, a lógica paraconsistente LP, com sua interpretação canônica, em que *g* serve para representar dialeteias, simplesmente autoriza o tratamento lógico de contradições verdadeiras, mas ela mesma não garante a existência de tais contradições (essa existência não seria uma consequência apenas da lógica). Para tanto, precisamos de uma motivação adicional, que estabeleça de algum

[15] Recorde que na introdução, colocamos um argumento em que inferimos uma sentença, β, qualquer a partir de uma contradição $\alpha \wedge \neg \alpha$. Para tanto, usamos, dentre outras coisas, a regra de silogismo disjuntivo (SD). LP bloqueia aquela derivação, uma vez que SD não vale em LP. Para conferir isso, basta tomarmos o mesmo contra modelo usado acima (i.e, α com o valor aglutinado e β com o valor falso).

[16] Para uma apresentação de LP, além da que está no artigo original de 1979, veja PRIEST, 2008, cap. 7.

modo, a existência de algumas de tais contradições. Há vários tipos de motivações que são discutidas na literatura sobre o dialeteísmo: situações envolvendo legislação inconsistente, descrições de transição de estados requerendo inconsistência, predicados vagos, paradoxos de autorreferência, dentre outros (cf. PRIEST, BERTO, WEBER, 2018; PRIEST, 2007a). Os paradoxos de autorreferência são colocados como os "maiores" e "mais impressionantes" argumentos em favor do dialeteísmo (cf. PRIEST, BERTO, WEBER, 2018, p. 14; PRIEST, 2006b, p. 83).[17] De fato, os paradoxos semânticos, como o Paradoxo do Mentiroso, são os mais frequentes na literatura dialeteísta, e são aqueles que ganharam maior influência na literatura quando se trata de motivar (e também de criticar) o dialeteísmo. Vamos focar aqui, então, no paradoxo do Mentiroso (ou simplesmente, Mentiroso) como um argumento em favor do dialeteísmo.[18]

O paradoxo do Mentiroso pode ser construído facilmente a partir de uma sentença que afirma a sua própria falsidade.

(λ) λ é falsa.

A questão que se coloca é: qual é o valor de verdade de λ? Por um lado, se λ for verdadeira, temos que ela é falsa. Por outro lado, se λ for falsa,

[17] Paradoxos de autorreferência englobam os paradoxos semânticos (que surgem através de conceitos como verdade, denotação, etc.) e os paradoxos conjuntistas (que surgem através de conceitos como de pertinência, cardinalidade, etc.). Com essa divisão, dentre os paradoxos semânticos, temos paradoxos como o do Mentiroso, e dentre os paradoxos conjuntistas, temos paradoxos como de Russell. Essa divisão é discutível, sobretudo depois dos trabalhos de Tarski que faz uso de teorias de conjuntos e lida com o paradoxo do Mentiroso. No contexto dialeteísta, ainda que apareça para fins didáticos a divisão entre paradoxos semânticos e paradoxos conjuntistas, Priest defende que tais paradoxos compartilham da mesma "estrutura essencial", chamada por ele de "*inclosure schema*" (cf. PRIEST, 2002, part. III).

[18] Vale notar que existe outro paradoxo de autorreferência, o paradoxo de Curry, que também gera problemas para o dialeteísmo. Em sua versão mais simples, a sentença de Curry é uma condicional C dizendo intuitivamente que "se C é verdadeira, então A", onde "A" é uma sentença qualquer. Com alguns poucos recursos lógicos e a definição de C, podemos provar A. Através desse paradoxo, a trivialidade pode ser alcançada diretamente, sem que passe por uma contradição. (Para uma apresentação introdutória do problema do paradoxo de Curry no contexto dialeteísta, ver READ, 2016, pp. 207-209). Para lidar com o paradoxo de Curry, Priest oferece um tratamento do condicional que restringe o princípio de absorção, empregado na derivação do paradoxo (Cf. PRIEST, 2006a, sec. 6.2). Há uma vasta literatura sobre os paradoxos do tipo Curry no contexto dialeteísta. Tendo em vista o escopo do artigo, não vamos entrar nessa discussão aqui. (Para uma discussão mais recente sobre paradoxos do tipo Curry no dialeteísmo, veja ARENHART & MELO, 2019b).

ela será verdadeira. Temos, portanto, que λ é verdadeira se e somente se é falsa. Levando em consideração que toda sentença é verdadeira ou falsa, λ deve ser verdadeira e falsa. Isso seria um problema para concepções tradicionais da verdade, nas quais vale alguma forma do LNC, que proíbe que sentenças sejam verdadeiras e falsas. Todavia, para aqueles que desconfiam da validade do LNC, ou para aqueles que estão convictos de que LNC pode falhar em alguns casos, pode muito bem ser o caso que, na conclusão do raciocínio em questão, temos uma aglutinação (*glut*) de valor de verdade. Em outras palavras, assumindo que o raciocínio é correto, temos uma contradição verdadeira (dialeteia), segundo os dialeteístas (cf. BEALL, 2009, cap. 1; PRIEST, 2006a, cap. 1).

Tendo isso em vista, alguém poderia tentar evitar essa conclusão justamente ao argumentar contra algum dos princípios que foram assumidos no argumento para se chegar a essa conclusão, aparentemente não tão fácil de se aceitar. Se a tentativa de caracterizar λ como verdadeira ou falsa, nos levou à contradição, pode ser que o problema resida justamente em assumir que há apenas esses valores para qualificarmos a sentença em questão. No lugar de aceitar que há contradições verdadeiras, alguém poderia sugerir que há sentenças que não são nem verdadeiras nem falsas; ou seja, assumir lacunas (*gaps*) de valores de verdade para caracterizar a sentença do Mentiroso. O problema dessa sugestão é que a partir dela surgem *formas fortalecidas do paradoxo*. De fato, com base nessa sugestão, podemos, por exemplo, estabelecer uma sentença que afirma que ela própria *não é verdadeira*:

(σ) σ não é verdadeira.

A partir de um raciocínio similar ao que fizemos para o caso de λ, podemos derivar a conclusão de que σ é verdadeira se e somente se não é verdadeira. Isso nos daria uma contradição novamente, de modo que a estratégia proposta parece incapaz de evitar inconsistências. Sendo relutante em afirmar uma contradição, alguém poderia pensar que deve haver uma outra maneira de caracterizar a sentença do Mentiroso consistentemente.

Talvez o Mentiroso esteja justamente mostrando que verdade e falsidade não são exaustivas. Para frustrar essa ideia, é apresentado um fenômeno conhecido como *vingança do Mentiroso*: independentemente da caracterização usada para solucionar o Mentiroso consistentemente, é possível usar os termos da própria caracterização para chegar à contradição. Para deixar mais claro o problema da vingança, tome a seguinte sentença ϵ.

(ϵ) ϵ é falsa ou *X*.

Aqui *X* poderia ser qualquer termo escolhido para caracterizar o Mentiroso. Com base em ϵ, temos uma espécie de fórmula da vingança: a partir de qualquer conceito colocado no lugar de *X,* é possível estabelecer uma sentença que, através de um raciocínio com bastante apelo intuitivo, leva à contradição.[19]

Note, por exemplo, que σ poderia ser colocada na forma geral de ϵ da seguinte maneira:

(σ) σ é falsa ou não verdadeira.

Com base na vingança do Mentiroso, no contexto dialeteísta, é argumentado que as teorias só podem adquirir a consistência ao custo de deixar uma noção importante inexpressível na linguagem (geralmente justamente aquela que foi colocada para resolver o paradoxo). Se quisermos uma linguagem com completude expressiva (i.e., poder expressar tudo que temos disponível na linguagem, inclusive os conceitos semânticos) a inconsistência será inevitável. Tendo isso em vista, como resultado da vingança é estabelecido o seguinte dilema, que o dialeteísta usa em seu favor, e contra abordagens consistentes ao paradoxo: completude expressiva ou consistência. Os dois disjuntos não podem ser mantidos, frente ao problema da vingança. Segundo dialeteístas, o fenômeno evidencia a resiliência do Mentiroso diante de tratamentos consistentes; ao garantir a consistência,

[19] Para uma discussão geral sobre o fenômeno da vingança, veja BEALL, 2007.

perdemos poder expressivo, e ao permitir poder expressivo sem restrições, perdemos a consistência.

Nesse sentido, Priest defende que há uma característica fundamental do Mentiroso que é preservada através das várias formulações possíveis do paradoxo, através de suas formas estendidas (ou reforçadas). Priest apresenta essa característica em termos daquilo que ele chama de "essência do Mentiroso". Para explicar isso, Priest divide as sentenças entre verdades legitimas (Priest chama de '*bona fide truths'*) e resto (entendido como o complemento da classe das verdades legítimas) e define a "essência do Mentiroso" nos seguintes termos:

> a essência do paradoxo do Mentiroso é uma construção particular retorcida que força uma sentença, se ela está nas verdades *bona fide*, a estar no resto (também); conversamente, se ela está no resto, ela está nas verdades *bona fid*e. (PRIEST, 2006a, p. 23).

No jargão lógico, o conjunto de verdades legítimas é identificado como o conjunto designado e resto é identificado como o conjunto não designado. Na linha de argumentação dialeteísta, a essência do Mentiroso captura o que há de fundamental nas diferentes formulações do paradoxo: a alternância da sentença paradoxal entre as classes de verdades legítimas e resto (independentemente de como for caracterizado a classe resto).

Priest usa essa noção de essência do Mentiroso para argumentar contra os tratamentos que assumem lacunas de valores de verdade. Considerando o paradoxo que surge a partir da sentença do Mentiroso simples, representada por λ acima, temos que as verdades legítimas são simplesmente as sentenças recebendo o valor verdadeiro, e o complemento coincide com o falso. Se tentarmos resolver o problema admitindo lacunas (ou seja, se as sentenças podem, além de ser verdadeiras ou falsas, ter uma lacuna), "ser falsa" acaba sendo apenas uma parte própria do resto e isso, segundo Priest (2006a, p. 23), não coloca o problema de maneira apropriada. Note o que está em jogo: ao assumirmos três valores de verdade, incluindo a lacuna, e tentar evitar a contradição do paradoxo afirmando que a sentença do Mentiroso é um caso de lacuna, a classe resto é ampliada (i.e., o resto fica sendo equivalente ao conjunto {falso, lacuna}). Isso evita a contradição

inicial, mas, ao mesmo tempo, é um problema para a caracterização do Mentiroso, porque a sentença do paradoxo passa a ficar fixa nessa classe, sem alterná-la para a classe de verdades legítimas. Com isso, temos a perda da essência do Mentiroso, que é precisamente essa alternância da sentença entre as duas classes (verdades legítimas e o resto). Ou seja, o paradoxo não está formulado adequadamente.

Todavia, isso não resolve realmente o problema dos paradoxos. Note que ao contemplar o novo cenário, em que a classe resto é ampliada para levar em conta a possibilidade das lacunas, podemos novamente obter um paradoxo. Isso é o que faz o movimento de vingança do Mentiroso. Em síntese, o que movimento da vingança faz é redescrever a classe resto, restabelecendo a essência do Mentiroso (identificada com a alternância entre as classes de verdades legítimas e seu complemento).

> Em um quadro em que o conjunto de sentenças é particionado em termos de uma tricotomia (verdadeiro, falso, e nem verdadeiro e nem falso) a natureza disjuntiva de 'esta sentença é falsa ou nem verdadeira e nem falsa' faz com que ela abranja todo o resto; i.e., o recém descrito complemento do conjunto das sentenças verdadeiras. E quanto adicionar mais valores? Se há uma quarta coisa que uma sentença pode ser, além de ser verdadeira, falsa, e nem verdadeira e nem falsa, podemos sempre tomar a noção *quarta coisa* e produzir outro Mentiroso reforçado (PRIEST, BERTO, WEBER, 2018, p. 17, grifo dos autores)

Tendo isso em vista, a vingança retrata o caráter resiliente do Mentiroso ao mostrar que a estabilidade provisória (de tomar uma posição em uma das classes) não funciona; a contradição sempre retorna quando redescrevemos a sentença paradoxal para levar em conta o novo cenário. Tudo o que precisamos fazer é garantir, na nova disposição de verdades legítimas e resto, que haja uma alternância entre as classes de verdades legítimas e seu complemento. Na linha dialeteísta, as versões estendidas (ou reforçadas) do Mentiroso evidenciam a inevitável característica do Mentiroso: a contradição. Segundo Priest (2006a, p. 24), "os paradoxos estendidos não são realmente novos paradoxos, mas meramente manifestações de um e do mesmo problema, apropriado a contextos diferentes".

O objetivo do dialeteísmo é evitar todas essas manobras que não funcionam, e fornecer uma resposta natural ao paradoxo: manter os princípios

intuitivos que levam à contradição e aceitá-la como um fato da vida (cf. PRIEST, 1979; PRIEST, BERTO, WEBER, 2018).

> De maneira geral, tais paradoxos como o Mentiroso fornecem alguma evidência para a tese dialeteísta que algumas contradições são demonstravelmente verdadeiras, no sentido que elas são implicadas por fatos manifestos sobre as línguas naturais e nossos processos de pensamentos. Os paradoxos do Mentiroso estendido como 'Esta sentença não é verdadeira' são expressáveis no português ordinário. Suas características paradoxais, como os dialeteístas salientam, são devidas justamente às características intuitivas da linguagem ordinária (PRIEST, BERTO, WEBER, 2018, p. 20).

A grande lição que os paradoxos semânticos nos dão, na perspectiva dialeteísta, é a de que há contradições verdadeiras e que, portanto, o dialeteísmo está correto. Diante do fracasso de teorias alternativas em fornecer uma resposta consistente que preserve o poder expressivo da linguagem natural, o dialeteísmo se oferece como a única alternativa razoável ao problema, acabando, de quebra, com a hegemonia da LNC na tradição ocidental.

4. A interpretação dialeteísta da contradição e da paraconsistência.

Como vimos, uma vez que contradições verdadeiras são assumidas, por conta do tratamento dado ao paradoxo do Mentiroso, o dialeteísmo precisa de uma lógica paraconsistente a fim de manter que existem contradições sem trivialidade. Porém, os dialeteístas são bastante seletivos quando se trata de especificar como opera essa lógica. É necessário que haja uma íntima conexão entre os cânones do dialeteísmo e o funcionamento da lógica paraconsistente subjacente. O sistema LP tem sido considerado como uma lógica apropriada para manter contradições verdadeiras sem trivialidade. Entretanto, nem toda interpretação de LP é adequada para o dialeteísmo. Como mencionamos acima, há interpretações não dialeteístas de LP — como o próprio Priest reconhece (cf. PRIEST, 2019b). Diante disso, devemos agora especificar mais claramente como se articulam LP e dialeteísmo para que LP resulte realmente um sistema adequado para o dialeteísmo. Para investigar isso, começamos pela própria noção de dialeteísmo. O dialeteísmo normalmente é definido como a tese de que algumas contradições, α e $\neg\alpha$,

são verdadeiras (veja alguns exemplos nas citações colocadas no início da seção 2). Note que nessa definição — assim como várias outras que aparecem frequentemente na literatura dialeteísta — o dialeteísmo está sendo apresentado a partir da noção sintática de contradição como um par de sentenças α e $\neg\alpha$. Para futuras referências, tomamos as seguintes noções:

C. Sint: sentenças da forma α e $\neg\alpha$ são contradições.
Dia: algumas contradições são verdadeiras.

Uma noção meramente sintática de contradição, como encapsulada em **C. Sint**, não é suficiente para o dialeteísmo. A interpretação dos conectivos, especialmente o conectivo da negação, é determinante para o funcionamento apropriado do dialeteísmo. Para que isso ocorra, LP precisa ser entendida de modo que o símbolo de negação envolvido seja interpretado (i) como um operador formador de contradição (seja um ofc) e (ii) não seja explosivo. Manter conjuntamente as condições (i) e (ii) é condição necessária para que o dialeteísmo faça sentido. De maneira mais explícita, para manter a coerência do dialeteísmo, sua lógica subjacente deve permitir contradições sem trivialidade. Fundamentalmente, é preciso deixar claro qual é o sentido de contradição relevante para o dialeteísmo. Essa é uma questão prévia a interpretação de LP. Isso vai ficar mais claro quando retomarmos a discussão da principal motivação para o dialeteísmo: os paradoxos semânticos.

Por questão de argumento, vamos assumir aqui o tratamento dialeteísta do paradoxo do Mentiroso. Como vimos, no dialeteísmo, o Mentiroso é considerado uma contradição verdadeira, independentemente da roupagem que ele possa receber, através de suas várias formulações; ele deverá manter a sua "essência". Priest (2006, p. 84) diz que o Mentiroso nos indica claramente que "verdade e falsidade se sobrepõem: para alguns αs, temos que α e $\neg\alpha$". Então, é possível termos $(\alpha \wedge \neg\alpha)$. Dessa forma, uma sentença do Mentiroso, como λ, será verdadeira e, sua negação, $\neg\lambda$, também será verdadeira. Note que, pela argumentação dialeteísta, que colocamos na

seção 3, podemos partir de qualquer sentença do Mentiroso, como λ e σ, e teremos o mesmo resultado: a sentença e sua negação serão verdadeiras.

Como vimos, por meio das tabelas de LP, o comportamento da negação permite que uma sentença, α, e sua negação, $\neg\alpha$, sejam verdadeiras. Mas, note que até então, o significado da negação está suspenso no ar (uma formulação sintática da contradição não nos garante a interpretação da contradição). Em seu influente artigo '*Paraconsistent Logic*?', Hartley Slater revela o sentido da negação de LP ao defender que ela tem o comportamento de um operador formador de subcontrárias — noção proveniente do quadrado de oposições de Aristóteles. Antes de continuar com o argumento de Slater, vamos lembrar rapidamente algumas noções do quadrado de oposições:[20]

> **Contrárias**: α e β são *contrárias* quando elas não podem ser ambas verdadeiras, mas podem ser ambas falsas.
> **Subcontrárias**: α e β são *subcontrárias* quando elas não podem ser ambas falsas, mas podem ser ambas verdadeiras.
> **Contraditórias**: α e β são *contraditórias* quando elas não podem ser ambas verdadeiras, nem ambas falsas.

A relação de contrariedade garante a exclusividade, na medida em que ambas sentenças não podem ser verdadeiras. A relação de subcontrariedade garante a exaustividade, visto que ambas sentenças não podem ser falsas. A relação de contraditoriedade garante tanto a exclusividade quanto a exaustividade.

Para defender que a negação de LP é formadora de subcontrariedade, Slater (1995) toma uma valoração *v* como sendo uma função das sentenças da linguagem no conjunto {1, 0, -1} (correspondente ao conjunto {*t, g, f*} que usamos na seção 2). Slater evidencia que em LP α e $\neg\alpha$ não são contraditórias (tomando essa noção conforme definida pelo quadrado de oposição). O argumento de Slater, grosso modo, é o seguinte: como a 'verdade

[20] Para uma visão geral sobre o quadrado de oposições, veja PARSONS, 2017.

de α' (i.e., $v(\alpha) \geq 0$) é equivalente a 'α', e, também, como a 'verdade de $\neg\alpha$' é equivalente a '$\neg\alpha$', temos que α e $\neg\alpha$ não são contraditórias, e sim subcontrárias. Em outros termos: tanto uma fórmula quanto sua negação podem possuir valor maior ou igual a 0; isso ocorre especificamente porque negar uma fórmula com valor 0 não nos leva a uma fórmula com valor -1. Ou seja, a negação não muda os valores de 0 ou 1 (as verdades legítimas, para usar terminologia da seção anterior) para -1 (a classe resto). Portanto, o operador de negação '¬' de LP não é um ofc (cf. SLATER, 1995, p. 451). Diferentemente do que ocorre em LP, a contraditória da 'verdade de α' (sentenças verdadeiras, ainda que algumas sejam falsas também) é a 'falsidade de α' (apenas falsa, sem que seja verdadeira também). Ou seja, apenas quando passamos, com o uso da negação, de sentenças com valores 0 ou 1 para sentenças com valor -1 (e reciprocamente) é que temos uma ofc.

Alguém poderia responder que a contraditoriedade é a interpretação dada para a negação clássica e que o fato da negação de LP não capturar o sentido clássico da negação não faz com que ela deixe de ser uma negação (cf. BÉZIAU, 2006). Mas essa defesa seria insuficiente em dois sentidos. Primeiramente, não está em questão (no presente contexto) se a negação dialeteísta é uma negação; o que está em questão é se ela é um ofc. Assim, garantir que um operador que não é ofc é uma negação ainda fica aquém do esperado. Em segundo lugar, e ainda mais importante para nossos propósitos, isso é insuficiente tendo em vista os propósitos do dialeteísmo. Priest considera que o argumento de Slater está errado e defende que a negação de LP captura justamente o sentido intuitivo ou pré-teórico de contraditoriedade. Ou seja, para Priest, a negação de LP é formadora de contradição, interpretada pela noção de contraditoriedade do quadrado de oposição de Aristóteles.

> Assim se α é qualquer enunciado, seja $\neg\alpha$ representando sua contraditória. (contraditórias, diferentemente das contrárias e subcontrárias são únicas — pelo menos até a equivalência lógica). Quais relações valem entre elas? A lógica tradicional e o senso comum são ambos muito claros sobre a relação mais importante: *devemos ter*

> *pelo menos um membro do par, mas não ambos*. Isso é precisamente o que distingue as contraditórias de suas primas próximas contrárias e subcontrárias (PRIEST, 2006b, p. 78, grifo nosso).

Agora temos um sentido prévio que a negação de LP deve retratar. A direção aqui é importante: a noção pré-teórica de contradição deve determinar o comportamento da teoria da negação, e não o contrário. Uma formulação sintática de $\alpha \wedge \neg \alpha$, usando uma negação arbitrária, não garante que o sentido de contradição pretendido está sendo assegurado; a interpretação da negação guiada por um sentido pré-teórico deve ser levada em conta para se garantir a adequação da teoria. Priest (2006b, p. 76) argumenta que não devemos confundir teoria da negação com a própria negação. Há várias teorias sobre a negação (clássica, paracompletas, paraconsistentes, etc.) que estabelecem como o símbolo de negação se comporta teoricamente. Porém, segundo Priest, há apenas uma negação ("objeto real") que uma determinada teoria da negação (objeto teórico) pode ou não retratar de maneira correta. Para Priest (2006b, p. 77), teorias da negação são teorias sobre a relação de contraditoriedade — esse é o sentido semântico pré-teórico que a teoria deve retratar, considerando os propósitos e aplicações do dialeteísmo. Para deixar mais claro, com base na citação de Priest acima, vamos definir contradição em termos semânticos da seguinte maneira:

> **C. Sem**: α e β são contraditórias precisamente quando temos pelo menos uma delas verdadeira, mas não ambas

A partir da noção de contraditoriedade (encapsulada em **C.Sem**), Priest diz que imediatamente temos as seguintes formulações de PTE e LNC:

> **PTE**: $\alpha \vee \neg \alpha$
> **LNC**: $\neg(\alpha \wedge \neg \alpha)$.

Mas note que, até então, essas fórmulas são apenas expressões sintáticas (para mais discussão sobre o sentido pré-teórico da negação segundo os dialeteístas, veja também a exposição em ARENHART e MELO,

2017, seção 2). Considerando que LP valida essas formulações sintáticas de LEM e LNC, Priest conclui que o símbolo de negação de LP é um ofc, com o sentido do quadrado de oposições. Essa é a resposta oficial de Priest para Slater (cf. PRIEST, 2006b, cap. 4; PRIEST, 2007b). Tendo em vista a resposta de Priest, o critério para conferir se uma dada teoria tem um símbolo de negação como um ofc é se ela valida as formulações sintáticas de PTE e LNC.

Com base nesse critério, Priest argumenta que lógicas paraconsistentes, como os sistemas C_n de da Costa, não são formadoras de contradição, justamente por violarem LNC — elas são apenas formadoras de subcontrárias. Na mesma linha, Priest argumenta que lógicas paracompletas também não garantem que a negação envolvida é formadora de contradição, na medida em que violam PTE. Para exemplificar isso, Priest toma as lógicas intuicionistas com sentenças que são interpretadas como sendo nem verdadeiras nem falsas. Essas lógicas não capturam a relação de contraditoriedade, e sim a relação mais fraca de contrariedade. Sobre isso, Priest diz:

> Se ¬ se comporta como sugerido, ela não é um operador formador de contradição de modo algum – meramente um operador formador de contrária. Isso pareceria particularmente claro se considerarmos a explicação intuicionista da negação. De acordo com ela, $\neg\alpha$ é verdadeira (= asserível) apenas se existe uma prova de que não há nenhuma prova de α. Isso obviamente é uma contrária de α (PRIEST, 2006b, p. 79).

Assim, uma mera formulação sintática da contradição, como um par α e $\neg\alpha$ (como em **C. Sint)**, não garante que ela captura o sentido pré-teórico da contradição (encapsulado em **C. Sem**). O símbolo de negação deve ter a interpretação apropriada, capturando o sentido pretendido de contradição.

Diferentemente das negações das lógicas paracompletas e paraconsistentes mencionadas acima, Priest argumenta que a negação de LP é um ofc, garantindo o sentido de **C.Sem**, visto que tanto $\alpha \vee \neg\alpha$ quanto $\neg(\alpha \wedge \neg\alpha)$ resultam válidas em LP (veja a seção 2). Mas, considerando que a negação de LP permite que α e $\neg\alpha$ sejam verdadeiras, tendo aglutinação de valores de verdade, como ela pode ser um ofc? Parece que a LNC deveria

falhar. Priest (2006b, p. 85) diz que o fato de termos algumas contradições serem verdadeiras, não impede que tenhamos uma negação como um ofc. Isso só mostra, segundo Priest, que a negação tem um conteúdo excedente (*surplus content*). O conteúdo excedente da negação é justamente o fato dela possibilitar α e $\neg\alpha$ serem verdadeiras. Com isso, segundo o dialeteísta, a negação de LP pode ser interpretada como um ofc e ainda assim, impedir a validade da explosão a partir de α e $\neg\alpha$.

5. A manobra de esticar a verdade alcança a contradição do Mentiroso?

Como vimos, Priest acredita ter conseguido assegurar os requisitos mínimos para que o dialeteísmo faça sentido, mantendo contradições na forma sintática, interpretadas precisamente como contraditoriedade no sentido pretendido, sem trivialidade. Todavia, consideramos que o argumento de Priest para garantir a interpretação de LP como um ofc não estabelece o ponto desejado (para detalhes, ver Arenhart e Melo 2018). O trajeto argumentativo de Priest para garantir que a noção de contradição empregada por ele é apropriada faz um apelo direto à forma sintática de determinadas proposições, deixando a definição de contradição segundo o quadrado de oposições intocada. Note: pelo que foi dito na seção anterior, Priest argumenta, contra Slater, que sua negação é sim um ofc pelo mero motivo de que PTE e LNC são válidas, em suas formas sintáticas. Mas validar as formulações sintáticas de PTE e LNC não é suficiente para determinar o sentido pré-teórico requerido como sendo a interpretação da negação. Como vimos acima, a noção pré-teórica deve determinar o comportamento da negação, e não o contrário. Ora, da mesma forma que $\alpha \wedge \neg\alpha$ não garante necessariamente a noção de contradição pretendida (seguindo relato de Priest sobre as lógicas paracompletas e paraconsistentes destacado na seção precedente), $\alpha \vee \neg\alpha$ e $\neg(\alpha \wedge \neg\alpha)$ também não determinam necessariamente os sentidos pré-teóricos requeridos. A chave para essa questão está justamente no vínculo que Priest tenta estabelecer entre o sentido pré-teórico da contradição (encapsulado em **C.Sem**) e a validação das expressões $\alpha \vee \neg\alpha$ e $\neg(\alpha \wedge \neg\alpha)$ em LP.

Considerando a noção pré-teórica de **C.Sem**, dada uma sentença, α, e sua negação, $\neg\alpha$, elas são contraditórias quando temos *pelo menos uma, mas não ambas*. Nesse sentido, $\alpha \vee \neg\alpha$ deve garantir a exaustividade (pela exigência de "pelo menos uma") e $\neg(\alpha \wedge \neg\alpha)$ deve garantir a exclusividade (pela exigência de "não ambas"). São essas noções pré-teóricas que devem ser asseguradas. Sem tais noções, o vínculo entre a noção intuitiva e a validação de $\alpha \vee \neg\alpha$ e $\neg(\alpha \wedge \neg\alpha)$ é quebrado, e o argumento de Priest para sustentar que a negação de LP é um ofc, com o sentido pretendido, falha. Especificamente, a negação de LP não garante a exclusividade, mesmo validando $\neg(\alpha \wedge \neg\alpha)$ e, com isso, uma expressão sintática (como **C.Sint**) em LP não representa o sentido de contradição pretendido.

O ponto reside no artifício para garantir LNC. A expressão $\neg(\alpha \wedge \neg\alpha)$ só resulta válida em LP justamente por conta da flexibilização do conjunto de valores designados, possibilitada pela manobra de se esticar a verdade (veja a seção 2). Sem esse artifício, que permite a aglutinação de valores de verdade, não seria possível validar as sentenças $\alpha \vee \neg\alpha$ e $\neg(\alpha \wedge \neg\alpha)$. O dialeteísta poderia protestar e responder que a aglutinação não se trata de um artifício *ad hoc*, colocado apenas para validar $\alpha \vee \neg\alpha$ e $\neg(\alpha \wedge \neg\alpha)$ e, com isso, garantir o sentido de **C.Sem**, já que a aglutinação resulta de uma razão independente; a saber, a própria derivação do Mentiroso. Todavia, isso não muda o fato da negação de LP não capturar a noção intuitiva de contradição encapsulada em **C.Sem**. Isso só evidencia que LP não consegue modelar (capturar) a contradição envolvida no Mentiroso, considerando as intuições pré-teóricas que residem na "essência do Mentiroso".

Vamos considerar o caso do Mentiroso, mas antes disso vamos recapitular, de maneira objetiva, o desvio de Priest pela sintaxe para tentar estabelecer interpretação da contradição com o sentido de **C.Sem**: a explicação da negação começa com a noção de contraditoriedade, encapsulada em **C.Sem**, e interpreta **C.Sint**. A partir daí, é extraído o sentido pré-teórico de PTE e LNC ("pelo menos uma, mas não ambas"). Através de considerações acerca do tratamento do Mentiroso e em particular da falha de abordagens consistentes em garantir completude expressiva, há a adição do

conteúdo excedente (a sobreposição de verdade e falsidade), proporcionando o esticamento da verdade que, por sua vez, invalida a explosão e valida PTE e LNC, tornando **Dia** compatível com **C.Sint**.

O problema é que o procedimento descrito acima enfraquece a semântica do símbolo de negação de LP. Nosso ponto aqui é simples: ao adicionar o conteúdo excedente como um novo valor de verdade, há uma interferência explícita no significado da negação. Em particular, a introdução da possibilidade de sentenças verdadeiras e falsas, possibilitado pelo artifício de se esticar a verdade, faz com que o sentido pré-teórico de LNC deixe de valer, e como consequência, mesmo que sintaticamente tenhamos uma versão de LNC ainda válida no sistema, a negação não é interpretada de maneira adequada como um ofc (com o sentido pretendido de **C.Sem**). Com isso, em LP, **C.Sint** não é interpretada como uma contradição do ponto de vista semântico ($\alpha \wedge \neg \alpha$ em LP não representa a contradição com o sentido de **C.Sem,** dado que esse já foi violado com a sobreposição de verdade e falsidade). Ao capturar o sentido pré-teórico de PTE e violar o sentido pré-teórico de LNC, o símbolo de negação de LP resulta como sendo um operador formador de subcontrárias, como Slater já havia advertido. Com o sentido alterado da negação, não podemos garantir o sentido pré-teórico pretendido (encapsulado em **C.Sem**). O vínculo entre **C.Sem** e a validação de PTE e LNC é quebrado em LP; validar $\alpha \vee \neg \alpha$ e $\neg(\alpha \wedge \neg \alpha)$ com a negação de LP não garante **C.Sem**. O diagnóstico para esse fato é: a estratégia de esticar a verdade (que torna algumas sentenças verdadeiras também sentenças falsas) permite tornar **Dia** compatível com **C.Sint**, mas não é capaz de alcançar o sentido pretendido da contradição. Temos, portanto, que a paraconsistência de LP é garantida ao custo de perdemos o sentido pretendido de contradição (encapsulada em **C.Sem**). Um resultado inaceitável para o dialeteísmo enquanto a tese que afirma que algumas contradições são verdadeiras, e não que algumas subcontrariedades são verdadeiras.

Esse ponto fica ainda mais claro quando consideramos o argumento de Priest contra a abordagem de lacunas (*gaps*). Ao assumir lacunas, PTE falha e,

com isso, uma negação em tal abordagem resulta como sendo um operador formador de contrárias. Quanto a isso, Priest diz explicitamente:

> Um operador formador de contraditórias será aquele que quando aplicado a uma sentença, α, cobre *todos* os casos em que α não é verdadeira. Dessa forma, ele é um operador $\neg$, tal que $\neg\alpha$ é verdadeira sse α não é verdadeira, i.e., falsa ou nem verdadeira nem falsa. (em português, tal operador deve ser algo como: *não é o caso que*). (PRIEST, 2006b, p. 79, grifos do autor).

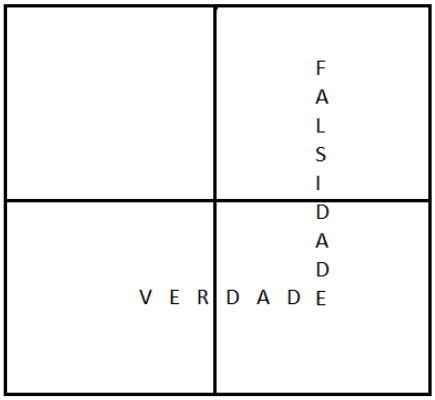
Como vimos, ao revelar a essência do Mentiroso, Priest divide as sentenças entre verdades legítimas (*bona fide truths*) e o seu complemento (resto). No caso de lacunas, a negação interpretada como um ofc deve alternar as sentenças verdadeiras para as sentenças não verdadeiras (falsas ou lacunas), e vice e versa. Mas, Priest (2006a, p. 23) argumenta que a abordagem de lacunas não faz isso, pois o 'falso' é interpretado como uma parte própria do resto. Uma negação é um ofc quando alterna entre a classe de verdades legítimas (designadas) e o seu complemento (não-designadas). Com isso, como vimos na seção anterior, na abordagem de lacunas, a sentença do Mentiroso permanece na classe resto, sem a essencial alternância entre as classes. O que o movimento da vingança faz é redescrever as classes de maneira exaustiva e exclusiva e, com isso, a negação é reinterpretada de modo a tornar a negação um ofc, alternando as sentenças entre as duas classes.

Todavia, algo muito semelhante acontece com a interpretação de LP. Como vimos na seção 2, em LP, há sentenças verdadeiras e somente verdadeiras (representadas por *t*), sentenças falsas e somente falsas (representadas por *f*) e sentenças aglutinadas, verdadeiras e falsas, (representadas por *g*). Como destacamos, a flexibilização do conjunto de valores designados, possibilitada quando se estica a verdade, é fundamental para invalidar a explosão e, ao mesmo tempo, validar PTE e LNC em LP (o que, segundo Priest, mostra que a negação de LP é paraconsistente e, ao mesmo tempo, formadora de contradição, garantindo, dessa forma, o sucesso do dialeteísmo). Assim, em LP, o conjunto designado (verdades legítimas) é ocupado pelas sentenças verdadeiras, ainda que algumas sejam falsas também, {*t*, *g*}, e o conjunto não-designado (resto) é ocupado pelas sentenças falsas, e somente falsas,

$\{f\}$. Uma negação interpretada como um ofc, com o sentido de **C.Sem** (que coincide com a essência do Mentiroso), deve alternar entre essas duas classes. Levando isso em conta, a condição de verdade para negação deve ser a seguinte:

> **ofc**: $\neg\alpha$ é verdadeira sse α não é nem verdadeira e nem aglutinada (i.e., apenas falsa).

Vamos recordar a citação de Priest acima: aplicar uma negação, interpretada como ofc, em uma sentença α deve cobrir *todos os casos em que* α *não é verdadeira.* Porém, uma negação com esse tipo de comportamento é explosiva. Sob essa interpretação, não há recursos, como o conteúdo excedente, capaz de sobrepor α e $\neg\alpha$, para impedir a trivialidade. Temos, portanto, a contradição (interpretada como **C.Sem**), ao custo de perdemos a paraconsistência. Um resultado inaceitável para o dialeteísmo, na medida em que põe em xeque a coerência do dialeteísmo. A figura da verdade exposta por Priest (2004, p. 27) pode deixar esse ponto um pouco mais claro.



No quadrante superior direito estão as sentenças falsas e somente falsas; no inferior esquerdo estão as sentenças verdadeiras e somente verdadeiras; no inferior direito estão as sentenças aglutinadas (verdadeiras e falsas); no quadrante superior esquerdo estão as lacunas (nem verdadeiras e nem falsas). Ao banir as lacunas, Priest diz que o quadrante superior esquerdo

não pode ser ocupado. O lado superior do quadrado representa, portanto, as sentenças que são apenas falsas. O lado inferior do quadrado representa a verdade esticada, que permite sentenças verdadeiras, mesmo que algumas delas também sejam falsas. A figura acima pode representar também a divisão, exaustiva e exclusiva, entre verdades legítimas (*bona fide truths*) e o seu complemento (resto); divisão necessária, como vimos, para representar a essência do Mentiroso (que captura a interpretação da negação como ofc). Temos, dessa forma, o lado inferior com as verdades legítimas (designadas) e o lado superior com o resto (não-designadas). Note que algo semelhante ao tratamento com lacunas acontece com o tratamento com aglutinações: o conjunto de sentenças aglutinadas (que estão no quadrante inferior esquerdo) é uma parte própria do conjunto das sentenças verdadeiras (que ocupa o lado inferior inteiro da figura).

A partir dessa divisão exaustiva e exclusiva, temos as sentenças verdadeiras, $\{t, g\}$, localizadas no lado inferior, e as sentenças falsas, $\{f\}$, localizadas no lado superior da figura. Aplicar a negação de LP (não explosiva) na sentença do Mentiroso, λ, não alterna entre as classes; λ e $\neg\lambda$, interpretadas como aglutinação, permanecem no lado inferior do quadrado, sem a essencial alternância entre a classe das verdades legítimas e o seu complemento. Assim, λ e $\neg\lambda$ não capturam a essência do Mentiroso; a negação nessa interpretação não é um ofc (com o sentido de **C.Sem**). De maneira análoga com a abordagem de lacunas, é possível reinterpretar a negação, permitindo que ela alterne as sentenças entre as duas classes, de modo a torná-la um ofc: $\neg\alpha$ é verdadeira ou aglutinada sse α é falsa (apenas falsa).

Essa negação mais forte exclui a sobreposição entre as classes exclusivas e exaustivas. Sem essa sobreposição, não podemos manter α e $\neg\alpha$ verdadeiras de modo a evitar a explosão. Considerando a noção de validade, α e $\neg\alpha$ não têm modelo e a trivialidade pode ser derivada. Para evitar isso, teríamos que esticar mais um pouco o conjunto das verdades legítimas (a classe designada), obtendo um novo recurso, algo como uma "super-aglutinação", capaz de sobrepor a classe das sentenças falsas com a classe das verdades legí-

timas. Com esse novo recurso, a trivialidade seria impedida na presença de α e $\neg\alpha$, mas, em contrapartida, perderíamos a divisão exaustiva e exclusiva entre a classe de verdades legitimas e o seu complemento (elas se sobrepõem por meio da super-aglutinação). Com a super-aglutinação, a classe resto (não-designada) é colapsada com a classe de verdade legítimas (designadas). Assim, a negação não é um ofc (com o sentido de **C.Sem**). A super-aglutinação alcança as sentenças falsas, permite a paraconsistência, com o custo de perder a necessária separação exaustiva e exclusiva das classes.

Com base na completude expressiva, o dialeteísta poderia colocar novamente a vingança a seu favor e redescrever uma nova classe resto, mantendo uma separação exaustiva e exclusiva entre essa nova classe e a classe de verdades legítimas. Isso recuperaria novamente a essência do Mentiroso e a negação poderia se comportar como um ofc (capturando o sentido de **C.Sem**). Todavia, como já era de se esperar, nesse novo cenário, não haveriam recursos para evitar a explosão a partir de α e $\neg\alpha$, com a negação reinterpretada. Novamente, para evitar a trivialidade, seria necessário esticar ainda mais o conjunto das verdades legítimas, obtendo um novo recurso, uma "super-super-aglutinação", capaz de sobrepor as sentenças da nova classe resto com as sentenças da classe das verdades legítimas. Isso poderia continuar para sempre como uma espécie de vingança contra o dialeteísta. Por um lado, se a negação de LP é interpretada como um ofc, ela perde a paraconsistência, não podendo impedir a trivialidade a partir da contradição $\alpha \wedge \neg\alpha$. Por outro lado, se ela é não explosiva, a negação não pode ser interpretada como um ofc e $\alpha \wedge \neg\alpha$ não representa uma contradição (com o sentido de **C.Sem**). Isso leva a um dilema difícil para o dialeteísta: manter contradições com trivialidade ou manter a paraconsistência sem contradição. Qualquer escolha determina o fracasso da visão que promete manter contradições verdadeiras, sem trivialidade.

Em última instância, independentemente de quantas vezes a classe das verdades legítimas for esticada, criando recursos para evitar a trivialidade, sempre haverá uma divisão entre verdades legítimas (designada) e o seu complemento resto (não-designada) capaz de preservar a essência do Mentiroso (i.e., o significado relevante de contradição, na perspetiva dialeteísta). O pro-

blema é que a negação de LP, capaz de evitar a trivialidade por meio de algum tipo de aglutinação, não alcança a contradição presente no Mentiroso. Com isso, independentemente de quantas vezes a classe das verdades legítimas for esticada ela nunca irá alcançar o Mentiroso (que depende da separação exaustiva e exclusiva entre a classe das verdades legítimas e seu complemento).

Tendo isso em vista, algo muito parecido com o que acontece com teorias que adotam lacunas, acontece com o dialeteísmo: enquanto que a teoria de lacunas não consegue garantir a exaustividade, o sentido pré-teórico de PTE, a teoria de aglutinação, assumida pelo dialeteísmo para lidar com contradições verdadeiras, não consegue garantir a exclusividade, sentido pré-teórico de LNC.[21] Com isso, temos que a negação de LP não é um ofc e, por conseguinte, LP não consegue representar o Mentiroso, uma das suas principais motivações e aplicações do dialeteísmo.

O que parece resultar claro dessas discussões é uma espécie de escolha que deve ser feita: por um lado, podemos fazer uma separação exclusiva e exaustiva dos valores de verdade. Se a negação descreve adequadamente essa separação, trocando os valores entre as duas classes, temos uma condição adequada para expressar tanto as contradições, quanto a essência do Mentiroso. O problema é que é exatamente nessas condições que a trivialidade resulta da negação que acompanha essa imagem. Por outro lado, podemos optar por uma visão da verdade como não exclusiva ou não exaustiva com a falsidade (ou até mesmo ambas), com uma negação correspondente, que mapeia aglutinações em aglutinações, e lacunas em lacunas. No caso de aceitarmos aglutinações, como o dialeteísta faz, temos o que intuitivamente seria o resultado do Mentiroso englobado no conjunto de valores de verdade. Isso permite que se descreva uma negação paraconsistente, que o dialeteísta necessita para evitar a trivialidade. Todavia, a imagem que resulta não é satisfatória em outros aspectos importantes também para o dialeteísta. Na medida em que a divisão entre verdade e falsidade não é exclusiva, aparentemente perdemos parte da essência do Mentiroso, conforme descrita pelo próprio dialeteísta. Do mesmo mo-

[21] Para um trabalho centrado na ideia de que o dialeteísmo não pode expressar a exclusão, ver ARENHART & MELO, 2019a.

do, a correspondente teoria da negação falha em descrever um operador formador de contradições. Não conseguimos ter ambos os quadros ao mesmo tempo sem gerar uma das dificuldades, perda da paraconsistência, ou perda do sentido pretendido de contradições.

6. Conclusão

Dialeteísmo é a visão de que algumas contradições são verdadeiras. Para manter tal visão sem trivialidade, é necessário assegurar a paraconsistência. Como foi destacado, a interpretação da paraconsistência deve desempenhar um papel crucial, tendo em vista a sua aplicação no dialeteísmo. Dizer que o dialeteísmo é a visão de que sentenças do tipo $\alpha \wedge \neg \alpha$ (ou α e $\neg \alpha$) são verdadeiras, sem oferecer uma interpretação adequada para o símbolo de negação, é insuficiente para o dialeteísmo. Além de ser considerado uma das mais evidentes motivações para o dialeteísmo, o paradoxo do Mentiroso revela o sentido de contradição adequado para o dialeteísmo. A Lógica do Paradoxo (LP) tem sido considerada por Priest e por outros dialeteístas como a lógica capaz de lidar com contradições verdadeiras, como a do Mentiroso, sem trivialidade. Para que o dialeteísmo faça sentido, a negação da lógica subjacente, no caso LP, precisa ser (i) um operador formador de contradição (ofc) e (ii) não permitir a explosão a partir de uma contradição (é crucial que (i) e (ii) sejam conjuntamente mantidas). O sentido pré-teórico da contradição, que é encontrado naquilo que Priest chama de "essência do Mentiroso", corresponde ao sentido da relação de contraditoriedade do quadrado tradicional de oposição. Assim, com base na noção de contradição (como contraditoriedade) um símbolo de negação interpretado como um ofc deve ser exaustivo e exclusivo.

Como vimos, o dialeteísmo é apresentado como uma abordagem que supera o problema da vingança do Mentiroso, tendo como vantagem a completude expressiva (capacidade de expressar todos os conceitos presentes na linguagem). Em um viés dialeteísta, a vingança do Mentiroso está apenas mostrando que há contradições verdadeiras (dialeteias) e que, portanto, o dialeteísmo é a visão correta. Para os dialeteístas, LP pode capturar a

contradição no Mentiroso, sem trivialidade, já que a sua negação é um ofc e ainda não é explosiva (i.e., LP cumpre as exigências (i) e (ii) acima). De fato, seria estranho evidenciar que há contradições verdadeiras por meio do Mentiroso e capturar um sentido de contradição que não corresponde ao Mentiroso.

Tomando então o sentido pré-teórico de contraditoriedade que corresponde à essência do Mentiroso, a negação como um ofc deve ser exaustiva e exclusiva. Todavia, considerando como o Mentiroso é retratado no dialeteísmo, temos, por exemplo, que a sentença, λ, e sua negação, $\neg\lambda$, resultam ambas verdadeiras. Em LP elas incorporam aglutinações de valor-verdade (*truth-value gluts*). Com isso, considerando o comportamento da negação em LP, temos que a negação é um operador formador de subcontrariedade, como Slater já havia apontado. Ao responder as acusações de Slater, Priest afirma que as formulações sintáticas de PTE, $\alpha \vee \neg \alpha$, e de LNC, $\neg(\alpha \wedge \neg \alpha)$ resultam válidas em LP e que, portanto, a negação de LP capturaria, respectivamente, a exaustividade e a exclusividade (componentes da noção de contradição como contraditoriedade).

Neste artigo demos um passo além no debate Priest-Slater sobre a natureza da negação e defendemos que os argumentos de Priest não funcionam. Como mostramos, Priest faz uma volta pela sintaxe para tentar garantir o conceito semântico de contradição. O dialeteísta tenta estabelecer um vínculo entre o sentido pré-teórico da contradição (como contraditoriedade) e a validação em LP das expressões sintáticas de PTE e LNC. Porém, como apontamos, esse vínculo só funciona caso os sentidos pré-teóricos de PTE e LNC — exaustão e exclusão, respectivamente — sejam mantidos. Como vimos, esticar a verdade permite aglutinação (sobreposição entre o verdadeiro e o falso), que garante a validação das expressões sintáticas de PTE e LNC e ainda permite que a explosão seja invalidada. Com isso, à primeira vista, LP cumpriria os requisitos (i) e (ii) apontados acima, garantindo o sucesso do dialeteísmo. Todavia, isso não acontece, tendo em vista a semântica de LP.

Da mesma maneira que uma formulação sintática da contradição, como $\alpha \wedge \neg \alpha$, não garante necessariamente que o sentido adequado da contradição está sendo retratado (como o próprio Priest argumenta), a validação das formulações sintáticas de PTE e LNC não garante necessariamente os seus sentidos pré-teóricos estão sendo retratados. Tudo depende da interpretação da semântica dos conectivos — neste caso, especialmente, a interpretação da negação. Como vimos, a validação das expressões sintáticas de PTE e LNC, bem como a invalidade da explosão só são garantidas por conta da flexibilização do conjunto de valores designados, possibilitada pelo fato de que se esticou a verdade. Todavia, como argumentamos, esse artifício enfraquece a semântica da negação de LP, quebrando o vínculo entre a validação das expressões sintáticas de PTE e LNC com o sentido pretendido da contradição. Portanto, a validação da expressão sintática de LNC em LP não garante a exclusividade. Com foi ressaltado, isso afeta o sentido da contradição, de modo que a expressão $\alpha \wedge \neg \alpha$ em LP não representa uma contradição no sentido pré-teórico requerido.

Como vimos, a essência do Mentiroso (que corresponde ao sentido pré-teórico de contradição adequado) depende de uma separação exaustiva e exclusiva da classe das verdades legítimas (conjunto designado) e da classe resto (conjunto não-designado). O problema é que quando esticamos a verdade, a exclusividade é perdida, o conjunto de sentenças aglutinadas é uma parte própria do conjunto das sentenças verdadeiras. Isso faz com que a negação de LP não seja um ofc (com o sentido de contradição requerido no dialeteísmo). Mas, como vimos, é possível redescrever, de maneira exaustiva e exclusiva, as classes das verdades legítimas e resto e reinterpretar a negação como um ofc. Porém, essa negação como ofc, ao garantir a exaustividade e a exclusividade, resulta explosiva. Isso poderia ser corrigido com o recurso de esticar a classe das verdades legítimas, a partir de um dispositivo de aglutinação, de modo que ela se sobreponha com a classe das sentenças falsas. Com isso, a explosão seria evitada, mas com o custo de perder a noção de

contradição com o sentido relevante (que depende da separação exaustiva e exclusiva das classes).

Esse processo entre esticar a verdade (para evitar a evitar a explosão) e a restruturação exaustiva e exclusiva das classes (para garantir a contradição) poderia ser reiterado indefinidamente. Porém, independentemente de quantas vezes a classe das verdades legítimas for esticada, ela nunca alcançará a contradição legítima exigida pela essência do Mentiroso. Esse processo impõe ao dialeteísta um dilema difícil sobre a interpretação da negação de LP: (i) ser operador formador de contradição ou (ii) não ser explosiva. Qualquer escolha levaria ao fracasso da doutrina que pretende sustentar que há contradições (no sentido requerido) sem trivialidade.

Referências

ARENHART, J. R. B. Interpreting Philosophical Interpretations of Paraconsistency. Talk at the *Ibero American Logic seminar*, 2020. http://www.ba-logic.com/seloi/ Acesso em 14 Jul. 2020.

ARENHART, J. R. B. & MELO, E. S. 'Classical Negation Strikes Back: Why Priest's Attack on Classical Negation Can't Succeed'. *Logica Universalis* (Print), v. 11, 2017, pp. 465 – 487.

_____'Dialetheists' Lies about the Liar'. *Principia: principia international journal of epistemology.* **22**, 2018 pp. 59-85. https://doi.org/10.5007/1808-1711.2018v22n1p59

_____ 'Dialetheism and the Inexpressibility of Exclusion'. *South American Journal of Logic*, v. 5, 2019a, pp. 213-229.

_____ '(De)motivating Gluts'. *Logique et Analyse*, v. 246, 2019b, pp. 115 – 154.

BEALL, J C. 'Introduction: at the Intersection of Truth and Falsity'. In. G., Priest, JC Beall, and B. Armour- Garb (eds.) *The law of non-contradiction. New philosophical essays*. Oxford: Oxford Un. Press. 2004, pp. 1–19

_____ 'Prolegomenon to future revenge'. In. Beall (ed.) *Revenge of the Liar. New Essays on the Paradox.* Oxford: Oxford Un. Press, 2007, pp. 1- 32

_____ *Spandrels of Truth*. Oxford, UK: Oxford University Press, 2009.

BÉZIAU, J.-Y. 'Paraconsistent Logics! A reply to Slater'. *Sorites*. 17. 2006, pp. 17-25.

CARNIELLI, W. & RODRIGUES, A. 'An epistemic approach to paraconsistency: A logic of evidence and truth'. *Synthese*, 196, 2017. pp. 3789–3813. https://doi.org/10.1007/s11229-017-1621-7.

CARNIELLI, W. & CONIGLIO, M. *Paraconsistent Logic: Consistency, Contradiction and Negation*, Dordrecht: Springer. 2016. doi:10.1007/978-3-319-33205-5

PARSONS, T., "The Traditional Square of Opposition", *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/square/>.

PRIEST, G. 'The logic of paradox'. *Journal of philosophical logic*. 8, 1979, pp. 219–241.

_____ 'Logic of paradox revisited'. *Journal of philosophical logic*. 13, 1984, pp. 153–179.

_____ 'Gaps and gluts: a reply to Parsons'. *Canadian Journal of Philosophy*. 25(1), 1995, pp. 57–66.

_____ *Beyond the Limits of Thought*, Cambridge: Cambridge University Press, 2nd expanded edition, Oxford: Oxford University Press, 2002.

_____ 'What's so bad about contradictions?' In: G., Priest, JC Beall, and B. Armour- Garb (eds.) *The law of non-contradiction. New philosophical essays*. Oxford: Oxford Un. Press. 2004, pp. 23-38.

_____ *In Contradiction: a study of the transconsistent*. 2nd edition. Oxford: Oxford Un. Press, 2006a.

_____ *Doubt truth to be a liar*. Oxford: Oxford Un. Press, 2006b.

_____ 'Paraconsistency and Dialetheism', In D. Gabbay & J. Woods (eds.). *Handbook of the History of Logic*, Amsterdam: Elsevier. viii. 2007a, ch. 3.

_____ 'Reply to Slater' In J-Y Beziau, W. Carnielli, D. Gabbay (eds.), H*andbook of Paraconsistency*. College Publications. London, U.K. 2007b, pp. 467 – 474.

_____ *An Introduction to Non-classical Logic*. Cambridge University Press, Cambridge, 2008.

_____ 'Crossing Boundaries'. In: Başkent C., Ferguson T. (eds*) Graham Priest on Dialetheism and Paraconsistency*. Outstanding Contributions to Logic, vol 18. Springer, Cham. 2019a, pp. 677-704.

_____ 'Some Comments and Replies'. In: Başkent C., Ferguson T. (eds*) Graham Priest on Dialetheism and Paraconsistency*. Outstanding Contributions to Logic, vol 18. Springer, Cham. 2019b, pp. 575-675.

PRIEST, G. ROUTLEY, R. (1989) 'The History of Paraconsistent Logic' In Priest, Routley; Norman (eds.) *Paraconsistent Logic: Essays on the Inconsistent.* München: Philosophia Verlag, 1989, cap. 1.

PRIEST, G., BERTO, F. and WEBER, Z., "Dialetheism", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/dialetheism/>.

READ, S. *Repensando a lógica: uma introdução à filosofia da lógica*. Trad. Abílio Rodrigues Filho. Belo Horizonte: Editora da UFMG, 2016.

SLATER, B. H. 'Paraconsistent logics?'. *Journal of Philosophical Logic*. 24. 1995, 451– 454.

EVIDÊNCIAS E DEPENDÊNCIA ONTOLÓGICA: SOBRE INTERPRETAÇÕES DE LÓGICAS PARACONSISTENTES[1]

Abílio Rodrigues[2]

Henrique Antunes[3]

RESUMO

Neste texto, apresentamos duas abordagens distintas, mas não necessariamente excludentes, da paraconsistência. A primeira trata das lógicas da evidência e verdade (*logics of evidence and truth*), nas quais contradições são interpretadas como evidência não conclusiva, simultânea e conflitante para um par de proposições A e $\neg A$. A segunda abordagem propõe que as diferenças entre a lógica clássica e alguns sistemas de lógica paraconsistente correspondem à distinção entre objetos ontologicamente independentes e ontologicamente dependentes, sendo esses últimos caracterizados como objetos cuja existência depende de nossos processos psicológicos e práticas linguísticas. Em ambas as alternativas, ilustramos como justificar a legitimidade das lógicas paraconsistentes sem endossar a tese de que a realidade, para ser descrita corretamente, requer proposições contraditórias.

Palavras-chave: Evidência. Dependência Ontológica. Contradição. Paraconsistência. Paracompletude.

ABSTRACT

In this paper we present two distinct, though not necessarily incompatible, approaches to paraconsistency. The first approach concerns the *logics of evidence and truth*, wherein contradictions are interpreted as conflicting non-conclusive evidence for a pair of propositions A & $\neg A$. The second approach proposes that the differences between classical logic and some systems of paraconsistent logic correspond to the distinction between *ontologically independent* and *ontologically dependent* objects, where the latter are characterized as objects whose existence depends on our psychological processes and linguistic practices. In either interpretation, we illustrate how to justify the

[1] O primeiro autor agradece o apoio do *CNPq* (Conselho Nacional de Desenvolvimento Científico e Tecnológico), projeto de pesquisa 311911/2018-8. Partes deste texto já foram publicadas em outros trabalhos dos autores, especialmente (Antunes 2019a) e (Carnielli e Rodrigues 2020).

[2] Professor associado do Departamento de Filosofia da Universidade Federal de Minas Gerais. E-mail: abilio.rodrigues@gmail.com. ORCID: 0000-0001-6639-9550.

[3] Pesquisador de Pós-Doutorado do Departamento de Filosofia da Universidade de Minas Gerais. E-mail: antunes.henrique@outlook.com. ORCID: 0000-0003-2089-8957.

legitimacy of paraconsistent logics without endorsing the thesis that reality itself requires contradictory propositions to be correctly described.

Keywords: Evidence. Ontological Dependence. Contradiction. Paraconsistency. Paracompleteness.

Introdução

Uma questão central na filosofia da lógica é se os princípios lógicos têm caráter ontológico, epistêmico ou linguístico, ou em outras palavras, como colocado por Popper (1963, pp. 206ss), se tais princípios dizem respeito à realidade, ao pensamento ou à linguagem. Esses três aspectos da lógica já podem ser implicitamente encontrados em Aristóteles, que formula na *Metafísica* três versões do princípio da não contradição, e cada uma delas pode ser relacionada, em algum sentido, com cada um desses aspectos[4].

Chateaubriand (2001) defende que o caráter fundamental da lógica não é linguístico, mas sim metafísico, o que significa que a lógica tem, por um lado, um aspecto ontológico, como uma "teoria sobre as características mais gerais e universais da realidade," e por outro lado um aspecto epistêmico, como "parte de uma teoria geral do conhecimento" (Chateaubriand 2001, pp. 16). Tugendhat e Wolf (1996, cap. 1) apresentam o problema a partir de uma perspectiva histórica, relacionando esses três aspectos a diferentes períodos da história da filosofia, respectivamente antiga e medieval, moderna e contemporânea. De fato, um traço característico da filosofia analítica é ter colocado a análise da linguagem como uma ferramenta fundamental para a discussão filosófica, mas isso não significa que os aspectos ontológicos e epistêmicos da lógica tenham sido esquecidos pela filosofia contemporânea. Esses três aspectos, na verdade, não são mutuamente exclusivos e podem ser combinados. A lógica pode ser aplicada para representar diferentes contextos argumentativos e dependendo das características de tais contextos um caráter ontológico, epistêmico ou linguístico pode ser predominante.

[4] Essa leitura de Aristóteles é encontrada, por exemplo, em Gottlieb (2013seção 1), Łukasiewicz (1910, 509ss), e discutida em Carnielli e Rodrigues (2016seção 3.2.4). As três versões do princípio da não contradição estão na *Metafísica* de Aristóteles, respectivamente, 1005b15, 1005b25, 1011b15.

Consequência clássica

No século XX, a lógica como campo de estudo foi dividida em diferentes disciplinas, mas é correto afirmar que o principal conceito da lógica é o de consequência lógica, de modo que a questão central da lógica é: *o que se segue de quê?*

O tratamento da consequência que chamamos de lógica clássica foi inaugurado por (Frege 1879) ao apresentar o sistema formal da *Begriffschrift*, que tem como fragmento um sistema dedutivo da lógica clássica de primeira ordem. O sistema de Frege é apresentado sintaticamente, mas a ideia subjacente é que as regras de inferência preservam verdade. Essa ideia é expressa de modo preciso pela definição semântica de consequência clássica:

(1) A segue de Γ ($\Gamma \vDash A$) se e somente se para todo modelo M, se as premissas em Γ são verdadeiras em M, então a conclusão A é verdadeira em M,

onde um modelo para uma linguagem L é um domínio D e uma função I que interpreta L sobre D. A definição (1) expressa a ideia pré-teórica de consequência lógica segundo a qual um argumento é válido quando é impossível que suas premissas sejam verdadeiras e sua conclusão falsa.

Entre os princípios lógicos que são válidos na lógica clássica estão a não contradição, o terceiro excluído e o princípio da explosão, expressos em uma linguagem proposicional respectivamente pelas fórmulas

(2) $\vdash \neg(A \wedge \neg A)$,

(3) $\vdash A \vee \neg A$,

(4) $A, \neg A \vdash B$.

Nas leituras clássicas, o princípio da não contradição (2) afirma que uma contradição, i.e. a conjunção de uma proposição A e sua negação $\neg A$, é sempre falsa, e o terceiro excluído (3) diz que dadas uma proposição A e sua negação $\neg A$, pelo menos uma delas é verdadeira. O princípio da explosão (4), também

chamado *ex contradictione quodlibet*, diz que qualquer proposição se segue de uma contradição.

A validade da explosão na lógica clássica é o resultado de considerar preservação de verdade não apenas como uma condição necessária, mas também como uma condição suficiente para consequência lógica. Como não há modelo M tal que A e $\neg A$ sejam verdadeiras em M, segue-se que para qualquer B, não há modelo M tal que A e $\neg A$ são verdadeiras e B é falsa em M. O papel do princípio da explosão na lógica clássica é normativo, pois obviamente ele nunca é de fato aplicado. Enquanto (2) diz que uma contradição é sempre falsa, (4), por outro lado, diz algo mais forte: afirma as consequências inaceitáveis que se deveriam enfrentar se uma contradição fosse tomada como verdadeira. É mais eficaz, portanto, que a não contradição para proibir contradições.

Como resultado da definição de consequência clássica como preservação da verdade, e considerando que verdade depende da realidade, é correto atribuir à lógica clássica um caráter predominantemente ontológico, como também realista, dado que o terceiro excluído é válido. Isso não significa que a lógica clássica não possa ser interpretada sem ingredientes ontológicos. Um sistema formal não é comprometido com uma única interpretação, ainda que tenha sido proposto com uma determinada interpretação pretendida. Mas a lógica clássica se encaixa perfeitamente bem com uma concepção de verdade realista e que rejeita contradições verdadeiras, e também com uma concepção ontológica da lógica. As lógicas não clássicas que serão discutidas a seguir permitem combinar a consequência clássica em termos de preservação da verdade com diferentes tratamentos da consequência lógica, recuperando a lógica clássica de tal modo que essas características da negação clássica são preservadas

Contextos contraditórios

Embora a lógica clássica não tolere contradições, é um fato que elas aparecem em vários contextos argumentativos. Bancos de dados muitas vezes contêm não apenas informações incompletas, mas também contraditórias. In-

vestigações científicas fornecem vários exemplos de situações reais em que contradições parecem ser inevitáveis[5]. Teorias matemáticas importantes permitiram provar resultados contraditórios no passado e, em virtude dos resultados de teoremas de incompletude de Gödel, não é possível saber com certeza se as teorias com as quais os matemáticos lidam hoje em dia estão livres de contradição. Personagens de ficção são algumas vezes descritas por seus autores como tendo propriedades contraditórias, seja intencional ou inadvertidamente. Ao contrário da lógica clássica, lógicas *paraconsistentes* são capazes de lidar com teorias e contextos argumentativos contraditórios.

Uma lógica é *paraconsistente* se, e somente se, tiver uma negação não explosiva, o que significa que o princípio da explosão (4) não é válido. Dada uma linguagem L e uma relação de consequência $\vdash$, um conjunto Γ de proposições de L é chamado *trivial* se, para todas as proposições A de L, $\Gamma \vdash A$. Um conjunto Γ é *contraditório* se, para alguma proposição A de L, $\Gamma \vdash A$ e $\Gamma \vdash \neg A$. Se o princípio da explosão for válido em $\vdash$, um conjunto Γ contraditório será também trivial. Mas se a relação de consequência $\vdash$ for paraconsistente, $\vdash$ permite conjuntos de proposições contraditórios mas não triviais.

Uma vez que existem contextos contraditórios e não triviais, a questão que se coloca é pela natureza das contradições que são admitidas em tais contextos, a qual está diretamente relacionada aos três aspectos da lógica mencionados anteriormente. O dialeteísmo é a tese segundo a qual existem dialetéias, que são sentenças ou proposições A tais que tanto A quanto sua negação $\neg A$ são verdadeiras (cf. Priest 1998, 2006; Priest, Berto, e Zach 2018). Se se admite que é a realidade que torna verdadeira uma proposição verdadeira, o dialeteísmo fornece uma resposta ontológica à pergunta pela natureza das contradições porque fica comprometido com o que chamamos de 'contradi-

[5] Existem teorias científicas bem-sucedidas em suas áreas de conhecimento que produzem contradições em determinadas circunstâncias específicas, ou quando combinadas com outras teorias também bem-sucedidas. Um exemplo, discutido em Carnielli e Rodrigues (2020seção 3.2.1), diz respeito ao surgimento da teoria da relatividade, que resolveu uma contradição entre a mecânica clássica e a teoria do campo eletromagnético. Sobre inconsistências em teorias científicas, ver e.g. (Colyvan 2008; da Costa e French 2003; da Costa e Krause 2014; Meheus 2002).

ções reais,' i.e., aspectos da realidade que para serem descritos corretamente exigem proposições contraditórias.[6]

Porém, essa não é a única maneira de justificar a rejeição do princípio da explosão pelas lógicas paraconsistentes. Contradições podem ser relacionadas a um conceito epistêmico, como justificativas não conclusivas e conflitantes, ou ter origem no pensamento ou no uso da linguagem. Em ambos os casos, proposições A e $\neg A$ que ocorrem em contextos contraditórios não são verdadeiras no sentido forte de serem tornadas verdadeiras por algum aspecto da realidade que seja intrinsecamente contraditório.

O objetivo deste texto é apresentar duas abordagens à paraconsistência que não estão comprometidas com contradições reais. A primeira, discutida na seção 2, trata das lógicas da evidência e verdade (*logics of evidence and truth*, (Carnielli e Rodrigues 2017; Rodrigues, Bueno-Soler, e Carnielli 2020)), nas quais contradições são interpretadas como evidência não conclusiva, simultânea e conflitante para A e $\neg A$. Essa abordagem é denominada 'epistêmica' porque rejeita contradições verdadeiras e enfatiza o ingrediente epistêmico presente na noção de evidência. A segunda abordagem, que é o foco da seção 3, propõe que as diferenças entre a lógica clássica e alguns sistemas de lógica paraconsistente correspondem à distinção entre objetos ontologicamente independentes e ontologicamente dependentes, sendo estes últimos caracterizados como objetos que dependem de nossos processos psicológicos e práticas linguísticas (cfe. Azzouni 2004). De acordo com essa abordagem, não há contradições acerca da realidade extra-linguística e extramental, muito embora contradições possam ser produzidas pela linguagem e

[6] Priest (2019, 588–89) afirma que a concepção de dialeteísmo por ele endossada não é comprometida com contradições na realidade, mas admite que essa leitura do dialeteísmo não é incomum. O que nós diríamos é que essa leitura do dialeteísmo de Priest é na verdade a predominante – ver, a título de exemplo, Batens (2002, 138), N. M. Kiefer (2019, 232), da Costa e O. Bueno (2007, 900), da Costa e French (2003, 106), da Costa e Krause (2014, 3041), Frápolli (1996, 438), Garfield (2004, 237), Mares (2004, 265), Meheus (2002, viii), H. Wansing (2016, 180), Steinberger (2016, footnote 2), entre outros. Além disso, também diferentemente do que lemos em (Priest 2019), pelo menos prima facie há suporte textual para uma leitura ontológica do dialeteísmo. Em Priest, Berto, e Zach (2018, Sect. 3.4), lemos que algumas dialetéias envolvem "contradições que afetam objetos concretos e o mundo empírico," e em Priest (2006, 159) que "devemos agora analisar contradições que surgem em domínios mais concretos, e especialmente no mundo empírico." Note-se, entretanto, que legitimidade de uma abordagem não ontológica da paraconsistência, que não considera verdadeiras as proposições contraditórias de contextos paraconsistentes, não depende dessa discussão acerca do dialeteísmo de Priest.

pelo pensamento – as quais dizem respeito ao que chamamos, seguindo (Azzouni 2004), de *objetos ontologicamente dependentes*.

Uma abordagem epistêmica

Contradições verdadeiras?

Um contra-exemplo ao princípio da explosão consiste em uma circunstância na qual um par de proposições A e $\neg A$ vale mas alguma proposição B não vale. A proposta do dialeteísmo, como vimos, é que existem circunstâncias em que proposições A e $\neg A$ são verdadeiras, muito embora o mundo não seja trivial (i.e., nem tudo o mais também seja verdadeiro), e isso seria suficiente como contraexemplo para o princípio da explosão. O dialeteísmo é sem dúvida uma visão não ortodoxa, discutida principalmente no contexto da filosofia da lógica. Matemáticos e investigadores das ciências empíricas não são muito interessados na questão de se existem ou não descrições da realidade contraditórias e verdadeiras, eles simplesmente pressupõem que não existem.

A formulação padrão do princípio da não contradição em termos de objetos e propriedades, que corresponde à versão ontológica da não contradição em Aristóteles [1005b15] e é amplamente aceita, é a seguinte:

(5) Não é possível um objeto a possuir e não possuir uma propriedade P ao mesmo tempo e no que diz respeito ao mesmo aspecto.

Uma contradição verdadeira portanto dependeria de um objeto a que possui não possui simultaneamente uma propriedade P ao mesmo tempo e no que diz respeito ao mesmo aspecto. Essas duas condições eliminam supostas contradições que dependem de perspectivas diferentes, como por exemplo estar frio e quente (i.e. não estar frio) para duas pessoas acostumadas com climas muito quentes ou muito frios, e também propriedades instanciadas em diferentes momentos, como por exemplo ser jovem e não ser jovem[7].

[7] Na tradição filosófica ocidental, os pensadores frequentemente mencionados como tendo defendido que a realidade é (em algum sentido) contraditória são Heráclito e Hegel. Mas em ambos as contradições estão relacionadas ao movimento e à mudança, e portanto é perfeita-

O apelo a propriedades pode ser dispensado. Basta recorrer à relação de satisfação utilizada por (Tarski 1956) para definir verdade, e uma contradição verdadeira $Pa \wedge \neg Pa$ dependeria então de um objeto a que satisfaz simultaneamente predicados P e $\neg P$. Assim, não é necessário apelar a entidades como fatos ou estados de coisas contraditórios, mas apenas a objetos contraditórios. De todo modo, a existência de dialetéias depende, na melhor hipótese, da existência de objetos contraditórios. Por essa razão, e em consonância com a leitura padrão do dialeteísmo, dizemos que o dialeteísmo fornece uma motivação ontológica para rejeição do princípio da explosão.

Evidência e verdade

O princípio da explosão pode ser rejeitado por razões puramente epistêmicas. A paraconsistência pode ser combinada com uma visão realista e não dialeteísta da verdade que simultaneamente aceita o terceiro excluído e rejeita contradições verdadeiras[8]. Segundo esse ponto de vista, as contradições que ocorrem em diversos contextos argumentativos não significam que proposições A e $\neg A$ sejam verdadeiras, mas apenas que existem evidências conflitantes para A e $\neg A$, e evidência para uma proposição A não implica a verdade de A (voltamos a esse ponto em breve).

Circunstâncias desse tipo ocorrem, por exemplo, nas ciências empíricas. Faz parte da rotina de físicos lidar com teorias que produzem contradições em circunstâncias críticas ou quando colocadas junto com outras teorias. O fato que quase todas as teorias científicas em algum ponto de seu desenvolvimento são contraditórias ou incompatíveis com outras teorias não é um problema para a racionalidade (cfe. Meheus 2002, vii). Há contradições que a prática científica não pode deixar de lidar, e contextos contraditórios temporá-

---

mente possível interpretá-los de tal modo que (5) não seja violado. A interpretação segundo a qual não há, nos fragmentos de Heráclito, contradições mas sim oposições e conflitos entre forças, estados e perspectivas é bem estabelecida na literatura. Quanto a Hegel, entretanto, dado que há trechos que mencionam explicitamente o princípio da não contradição, não fica claro à primeira vista que as 'contradições hegelianas' não são de fato violações do princípio da não contradição tal como formulado em (5) acima. Para um panorama sobre como contradições foram tratadas na história da filosofia, incluindo Heráclito e Hegel, ver (Rodrigues 2017).

[8] Sobre como conciliar uma visão realista da lógica com lógicas não clássicas, inclusive a paraconsistente, ver (Frade e Rodrigues 2020).

rios são um ingrediente essencial da prática científica (cfe. Nickles 2002, 2). Contradições podem ter origem em limitações do nosso aparato cognitivo, falhas em instrumentos de medida, estágios no desenvolvimento das teorias, ou simplesmente erros que podem ou não ser percebidos e corrigidos. Em todos esses casos, contradições são relacionadas em primeiro lugar ao pensamento e ao processo de aquisição de conhecimento, e tais contradições são chamadas de epistêmicas (cf. Carnielli e Rodrigues 2020, seção 3.2).

Considere um contexto argumentivo, como os mencionados acima, em que um par de proposições A e $\neg A$ são admitidas. A questão, então, é a seguinte: qual é a propriedade que vamos atribuir a A e $\neg A$ tal que seria possível uma proposição ter essa propriedade mesmo sem ser verdadeira? Tal propriedade tem de ser mais fraca que verdade, no sentido que uma proposição poderia ter essa propriedade sem ser verdadeira.

Evidência para uma proposição A é explicada informalmente em (Carnielli e Rodrigues 2017) como 'razões para acreditar em A.' Tais razões são objetivas, podem não ser conclusivas (evidência para A não implica a verdade de A), e não requerem crença por parte de um agente (evidência para A não implica que algum agente acredita em A). Evidência pode também ser explicada como um conteúdo semântico mais uma justificativa, que pode não ser conclusiva ou mesmo errada (cf. Carnielli e Rodrigues 2020; Rodrigues, Bueno-Soler, e Carnielli 2020). O ingrediente epistêmico da noção de evidência reside nessa justificativa que pode, no fim das contas, não ser uma justificativa efetiva. A noção de evidência, portanto, é uma propriedade mais fraca que verdade e fornece uma explicação intuitiva para a paraconsistência sem recorrer à contradições verdadeiras.

Informações contraditórias

Outra abordagem não dialeteísta da paraconsistência, no sentido de admitir contradições sem considerá-las verdadeiras, foi proposta nos anos 1970 por Belnap (1977b, 1977a) e Dunn (1976). A motivação de Belnap ao propor uma semântica de quatro valores para a lógica *FDE* (*logic of first-degree entailment*) era fornecer uma lógica a ser usada por um computador

que recebe informações não confiáveis de diferentes fontes. Essas informações podem ser incompletas ou contraditórias e são possíveis quatro cenários, representados pelos valores semânticos *Verdadeiro*, *Falso*, *Nenhum* e *Ambos*. Os valores *Verdadeiro* e *Falso* não são pensados como a verdade e a falsidade clássicas. Eles são explicados em termos de um computador que recebe a informação ('a computer being told') de que uma proposição é verdadeira ou falsa (Belnap 1977b, pp. 34–35). Esses quatro valores significam, respectivamente, que o computador recebe apenas a informação de que uma proposição A é verdadeira ('just told true'), apenas a informação de que A é falsa ('just told false'), não recebe nehuma informação sobre A ('told neither true nor false'), e recebe a informação de que A é verdadeira e a informação de que ela também é falsa ('told both true and false'). Belnap (1977b, pp. 47) afirma explicitamente que as proposições são verdadeiras ou falsas independentemente da informação que foi inserida no computador. Um computador pode receber a informação de que A é verdadeira (resp. falsa), mesmo que A seja falsa (resp. verdadeira). Assim como evidência, a noção de um computador receber a informação ('being told') de que uma proposição é verdadeira ou falsa é mais fraca que a noção de verdade, pois o computador pode receber a informação de que A é verdadeira mesmo que A seja falsa.

Lógicas da evidência e verdade

O operador de classicalidade

da Costa (1963, 1974), na sua hierarquia de lógicas C_n, propôs um 'operador de bom comportamento,' definido em C_1 em termos de não contradição,

(6) $A^\circ := \neg(A \wedge \neg A)$

que divide as proposições de uma dada linguagem em dois grupos: aquelas para as quais contradições são toleradas e as que contradições não são toleradas.

Pela definição (6) acima, uma proposição A é 'bem comportada,' expresso por A°, quando o princípio da não contradição vale para A.

As *LFI*s (*logics of formal inconsistency*) são um desenvolvimento da proposta de da Costa em que um operador de consistência $\circ$ é concebido de tal modo que $\circ A$ pode não ser equivalente a $\neg(A \wedge \neg A)$ (Carnielli e Coniglio 2016; Carnielli, Coniglio, e João Marcos 2007; Carnielli, Coniglio, e Rodrigues 2019). Nas *LFI*s, $\circ$ recupera o princípio de explosão para proposições no seu escopo:

(7) $A, \neg A \nvdash B$, ao passo que $\circ A, A, \neg A \vdash B$.

Embora na maior parte das *LFI*s investigadas na literatura a função do operador $\circ$ esteja circunscrita a recuperar o princípio de explosão (e inferências que dele se seguem), ele também pode desempenhar um papel mais geral em lógicas que, além da explosão, também invalidam o terceiro excluído. Em outras palavras, além de satisfazer (7), $\circ$ também permite recuperar $A \vee \neg A$:

(8) $\nvdash A \vee \neg A$, ao passo que $\circ A \vdash A \vee \neg A$.

As *LFIU*s (*logics of formal inconsistency and undeterminedness*) são lógicas nas quais $\circ$ tem ambas as propriedades (7) e (8), e nesse caso ele costuma ser chamado *operador de classicalidade* (e.g. Carnielli e Rodrigues 2017; Omori e Sano 2014; Rodrigues, Bueno-Soler, e Carnielli 2020; Szmuc 2016)[9].

As lógicas BLE e LET$_J$

Em (Carnielli e Rodrigues 2017), duas lógicas paraconsistentes são propostas: *BLE* (*basic logic of evidence*), que é a lógica *N4* interpretada

[9] Tal como ocorre no caso das *LFI*s que validam o terceiro excluído irrestritamente, lógicas que são paracompletas (i.e., nas quais o terceiro excluído não é válido), mas não paraconsistentes também podem incluir em sua linguagem um operador $\circ$ que satifaça *apenas* (8). Nesse caso, a lógica resultante é chamada *logic of formal undeterminedness* (*LFU*). Note que embora nem toda *LFI* ou *LFU* seja também uma *LFIU*, toda *LFIU* é tanto uma *LFI* quanto uma *LFIU*.

em termos de preservação de evidência[10], e LET_J (*logic of evidence and truth based on intuitionistic positive logic)*. LET_J é uma *LFIU* que estende *BLE* e na qual o operador ∘ permite lidar simultaneamente com contextos de dois tipos, um sujeito à lógica clássica, em as inferências preservam verdade, e outro sujeito a *BLE* em que as inferências preservam evidência.

A lógica *BLE* é definida em uma linguagem proposicional com os conectivos ¬, ∧, ∨ e → pelas seguintes regras de dedução natural:

$$\frac{A \quad B}{A \wedge B}\ \wedge I \qquad \frac{A \wedge B}{A}\ \wedge E \quad \frac{A \wedge B}{B}$$

$$\frac{A}{A \vee B}\ \vee I \quad \frac{B}{A \vee B} \qquad \frac{A \vee B \quad \begin{array}{c}[A]\\ \vdots \\ C\end{array} \quad \begin{array}{c}[B]\\ \vdots \\ C\end{array}}{C}\ \vee E$$

$$\frac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \to B}\ \to I \qquad \frac{A \to B \quad A}{B}\ \to E$$

$$\frac{\neg A}{\neg(A \wedge B)}\ \neg \wedge I \quad \frac{\neg B}{\neg(A \wedge B)} \qquad \frac{\neg(A \wedge B) \quad \begin{array}{c}[\neg A]\\ \vdots \\ C\end{array} \quad \begin{array}{c}[\neg B]\\ \vdots \\ C\end{array}}{C}\ \neg \wedge E$$

$$\frac{\neg A \quad \neg B}{\neg(A \vee B)}\ \neg \vee I \qquad \frac{\neg(A \vee B)}{\neg A}\ \neg \vee E \quad \frac{\neg(A \vee B)}{\neg B}$$

$$\frac{A \quad \neg B}{\neg(A \to B)}\ \neg \to I \qquad \frac{\neg(A \to B)}{A}\ \neg \to E \quad \frac{\neg(A \to B)}{\neg B}$$

$$\frac{A}{\neg\neg A}\ DN \quad \frac{\neg\neg A}{A}$$

As regras de *BLE* podem ser divididas em três grupos: (a) regras de introdução e eliminação de ∧,∨ e →; (b) regras de introdução e eliminação de ∧,∨ e → negadas; (c) regras de introdução e eliminação da dupla negação.

[10] Sobre *BLE* e *N4*, ver seção 5.3 e nota de rodapé 21 de (Carnielli e Rodrigues 2017).

A dupla negação não é suficiente para obter a lógica clássica porque em *BLE* uma fórmula $\neg\neg A$ somente pode ser introduzida a partir de uma premissa A. Isso ocorre porque não há regra de introdução da negação em *BLE*, i.e. a inferência abaixo não é válida:

(9) $A \rightarrow B, A \rightarrow \neg B \vdash \neg A$.

A implicação de *BLE* é não relevante, o que significa que vale o princípio conhecido como *ex quodlibet verum*, que corresponde à fórmula

(10) $\vdash A \rightarrow (B \rightarrow A)$.

O que (10) diz é que sendo A o caso, qualquer proposição implica A. Portanto, um contraexemplo para (9) é dado por uma circunstância em que há evidência conflitante para B e $\neg B$ mas nenhuma evidência para $\neg A$. A equivalência entre A e $\neg\neg A$ diz que 'evidência de que é falso que A é falso' é o mesmo que 'evidência de que A é verdadeiro.'

Os grupos (a) e (b) expressam, respectivamente, evidência positiva e negativa para conjunção, disjunção e implicação. Regras positivas e negativas independentes são necessárias porque evidência positiva e negativa são independentes e não complementares. Por essa razão, os quatro cenários abaixo são possíveis:

i. Apenas evidência que A é verdadeira: A vale, $\neg A$ não vale;
ii. Apenas evidência que A é falsa: $\neg A$ vale, A não vale;
iii. Ausência de evidência: nem A nem $\neg A$ vale;
iv. Evidência conflitante: ambas A e $\neg A$ valem.

A lógica LET_J é definida em uma linguagem que estende a linguagem de *BLE* com o operador de classicalidade $\circ$ e pelo acréscimo das seguintes regras ao sistema de dedução natural de *BLE*:

$$\frac{\circ A \quad A \quad \neg A}{B}\ EXP^{\circ} \qquad \frac{\circ A}{A \vee \neg A}\ PEM^{\circ}$$

As regras EXP° e PEM° recuperam a lógica clássica para as proposições no escopo de $\circ$, ou formadas com tais proposições e os conectivos $\neg,\wedge,\vee$, e $\rightarrow$ (Carnielli e Rodrigues 2017, Fact 18). A lógica LET_J é capaz de expressar os seguintes seis cenários:

- Quando $\circ A$ não vale, i.e. não há evidência conclusiva, temos os quatro cenários i-iv acima.
- Quando $\circ A$ vale, i.e. há evidência conclusiva, dois cenários são acrescentados:

  v. A vale: evidência conclusiva para A, i.e. A é verdadeira;

  vi. $\neg A$ vale: evidência conclusiva para $\neg A$, i.e. A é falsa.

LET_J trata evidência conclusiva como verdade porque considera que evidência conclusiva é sujeita à lógica clássica. LET_J pode também ser interpretada em termos de informação positiva ou negativa, que pode ser confiável ou não confiável. Nesse caso, os cenários i-iv são interpretados como informação não confiável e correspondem aos quatro valores da lógica *FDE* que vimos na seção 2.3, e os cenários v e vi são interpretados como informação confiável, e esta é considerada sujeita à lógica clássica (cfe. Antunes et al. 2020).

A lógica LET_J expressa em um mesmo sistema formal duas noções de consequência, uma que preserva verdade (proposições sujeitas à lógica clássica) e outra que preserva evidência. O operador $\circ$ pode ser considerado um comutador de contexto que muda o contexto de determinadas proposições, sujeitas à lógica clássica. LET_J, portanto, é capaz de expressar contextos multilógicos e é alinhada a uma concepção pluralista da lógica.

Segundo a interpretação pretendida, LET_J expressa uma abordagem da paraconsistência diametralmente oposta ao *dialeteísmo*. Se uma proposição A é verdadeira, i.e. se há evidência conclusiva para A, isso é expressado em LET_J pela conjunção $A \wedge \circ A$, e a falsidade de A, i.e. a verdade de $\neg A$, é ex-

pressada por $\neg A \wedge \circ A$. Assim, se A e $\neg A$ forem simultaneamente verdadeiras, temos $A \wedge \neg A \wedge \circ A$, e trivialidade se segue pela regra EXP°. De acordo com a interpretação pretendida de LET_J, portanto, contradições verdadeiras não são toleradas porque resultam em trivialidade, assim como na lógica clássica.

Uma abordagem nominalista

Como acabamos de ver, a abordagem epistêmica da paraconsistência fornece motivações bastante sólidas para rejeitar o princípio da explosão sem que seja necessário endossar a tese dialeteísta de que existem contradições verdadeiras. Vimos também que de acordo com essa interpretação não há um conflito real entre a lógica paraconsistente e a lógica clássica, uma vez enquanto a primeira fornece um tratamento da consequência lógica como preservação de verdade, a segunda trata da preservação de evidência (ou informação), que é uma noção reconhecidamente mais fraca.

Nesta seção, apresentaremos uma abordagem alternativa à paraconsistência segundo a qual as diferenças entre a lógicas clássica e alguns sistemas de lógica paraconsistente repousam em determinadas distinções ontológicas fundamentais entre dois tipos de objeto, entidades *ontologicamente independentes* e *ontologicamente dependentes* – uma distinção proposta por Jody (Azzouni 2004) no contexto do debate ontológico em filosofia da matemática[11]. A ideia central é que embora a lógica clássica forneça uma teoria satisfatória acerca de nosso raciocínio sobre domínios constituídos por objetos do primeiro tipo, determinados sistemas de lógica paraconsistente resultam ser mais adequados para raciocinar acerca de objetos do segundo.

Em virtude do fato de a distinção entre objetos ontologicamente dependentes e independentes ter sido proposta por Azzouni para sustentar a tese nominalista de que objetos matemáticos não existem, chamaremos essa segun-

[11] As ideias a serem discutidas aqui foram originalmente apresentadas em (Antunes 2018, 2019a).

da abordagem de *nominalista*, muito embora não trataremos dos problemas relacionados à ontologia da matemática neste artigo[12].

Diferentemente da interpretação epistêmica da paraconsistência, a abordagem nominalista não afirma que contradições não possam ser verdadeiras em absoluto, mas apenas que elas não podem ser verdadeiras sobre a realidade extra-linguística e extra-mental. Nesse sentido, contradições verdadeiras são vistas como um *mero subproduto* da maneira como a linguagem e o pensamento operam e não possuem nenhum correspondente na realidade. Há que se notar, contudo, que as duas abordagens não são necessariamente incompatíveis, pois enquanto a abordagem epistêmica trata a noção de verdade em um sentido estritamente realista (i.e., verdade como correspondência), a abordagem nominalista considera também a possibilidade de enunciados serem verdadeiros ou falsos em virtude de determinadas práticas linguísticas ou estipulações. Como veremos agora, é exclusivamente com relação a verdades desse segundo tipo que a abordagem nominalista admite a possibilidade de algumas contradições serem verdadeiras.

Objetos ontologicamente dependentes vs. independentes

Grosso modo, objetos ontologicamente dependentes são aqueles cujas propriedades dependem inteiramente dos processos psicológicos ou das práticas linguísticas de um indivíduo ou grupo de indivíduos, ao passo que objetos ontologicamente independentes são entidades cujas propriedades não dependem de tais processos ou práticas, mas da maneira como a realidade extra-linguística e mental (i.e., o mundo) se configura. Como exemplos paradigmáticos de entidades do primeiro tipo, temos figuras sonhadas ou imaginadas, alucinações e personagens de ficção; e como exemplos paradigmáticos de entidades do segundo tipo, temos objetos macrofísicos, tais como pessoas, animais, planetas etc.

[12] Ao invés de lidar com os objetos postulados pela matemática, como números, conjuntos, funções etc., consideramos apenas exemplos que envolvem personagens de ficção. A razão é que de acordo a proposta de Azzouni, tanto objetos matemáticos quanto ficcionais são ontologicamente dependentes, o que, segundo ele, significa que eles não existem realmente. Para uma discussão acerca das similaridades entre objetos matemáticos e ficcionais, ver (Azzouni 2015).

Embora esses dois tipos de objeto sejam fundamentalmente distintos, asserimos ordinariamente enunciados que são aparentemente verdadeiros ou falsos acerca de ambos, sem que as sentenças asseridas revelem qualquer diferença entre eles, seja em sua forma ou através de algum vocábulo específico da linguagem natural. Dizemos, por exemplo, que Dom Quixote inspirou muitas outras personagens famosas assim como que Miguel de Cervantes inspirou muitos outros autores da literatura mundial; ou que Dr. Jekyll e Mr. Hyde são a mesma personagem assim como que Fernando Pessoa e Alberto Caeiro são a mesma pessoa.

A despeito de não dispormos de meios linguísticos distintivos que tornem explícito quando estamos falando de entidades ontologicamente dependentes ou independentes, é natural supor que não raciocinamos acerca de ambos os tipos de objeto exatamente da mesma maneira – ou, mais precisamente, que eles não estejam sujeitos às mesmas relações de consequência lógica. Com efeito, personagens de ficção não parecem atender a determinados princípios característicos da lógica clássica, como o terceiro excluído ou a explosão, uma vez que elas são frequentemente caracterizadas de maneira incompleta e, em alguns casos especiais, até mesmo contraditória. Por outro lado, entidades ontologicamente independentes representam os candidatos mais naturais de objetos sobre os quais raciocinamos classicamente.

A tese de que existem enunciados acerca de entidades ontologicamente dependentes que não satisfazem o terceiro excluído não apresenta maiores dificuldades. Afinal, obras literárias raramente fornecem uma descrição exaustivamente completa de todas as suas personagens. Por outro lado, a ideia de que entidades ontologicamente dependentes sejam suscetíveis a descrições contraditórias é um tanto menos plausível. No entanto, é fácil fornecer exemplos que mostram justamente isso, como a seguinte história apresentada por Everett (2005, pp. 633–34):

> Quando chegou à terra da Dialetéia, Jane conheceu Jules e Jim. Isso a deixou confusa, uma vez que Jules e Jim eram e não eram pessoas distintas. No início, ela teve dificuldades para saber como deveria interagir com eles. Por exemplo, como Jules era e não era Jim, se Jim viesse à casa de Jane para tomar chá, Jules viria e não viria junto com ele. E nessas ocasiões ela não sabia quantos biscoitos servir.

> Jane se deu conta afinal o que fazer. Ela deveria comprar e não comprar mais biscoitos sempre que Jim viesse à sua casa. Depois disso, tudo ficou bem. (*nossa tradução*)

De acordo com essa história, Jim e Jules são e não são a mesma pessoa, o que implica que *as personagens* de Jim e Jules também são e não são a mesma[13]. Portanto, temos aqui um exemplo de objetos ontologicamente dependentes, viz., as personagens em questão, que satisfazem descrições contraditórias. E uma vez que o princípio da explosão proíbe toda e qualquer contradição, uma lógica adequada para raciocinar acerca de objetos desse tipo tem de ser paraconsistente.

Não pretendemos argumentar exaustivamente aqui a favor da tese que alguns objetos ontologicamente dependentes possuem propriedades contraditórias, ou que eles são suscetíveis a descrições contraditórias. No entanto, exemplos como o apresentado acima representam uma motivação inicial bastante forte a favor da tese de que enquanto a lógica clássica é correta para modelar nossos raciocínios acerca de objetos ontologicamente independentes, não se pode dizer o mesmo acerca de entidades ontologicamente dependentes. Afinal, em se tratando de objetos cujas propriedades foram todas estipuladas, não parece haver razões pelas quais algumas dessas estipulações não pudessem ser contraditórias, como no caso do fornecido por Everett. Assim, é possível distinguir entre as seguintes proposições:

(11) Não há objetos contraditórios; e

(12) Não há objetos ontologicamente independentes contraditórios,

onde (11) expressa que nenhum objeto, seja ele ontologicamente independente ou não, é passível de descrições contraditórias, ao passo que (12) restringe essa exigência apenas a objetos que existem independentemente de nós.

A lógica LFI1

[13] Parece ser indisputável que o único critério para determinar se duas personagens de uma mesma obra de ficção são ou não a mesma é se os indivíduos aos quais elas correspondem na história contada na obra são ou não idênticos. Afinal, que outra razão poderia existir para dizer que as personagens de Dr. Jekyll e de Mr. Hyde são a mesma, senão o fato de que esses indivíduos são, na verdade, a mesma pessoa de acordo com o livro de R. L. Stevenson?

O exemplo acima sugere que enquanto devemos subscrever a (12), (11) é um princípio demasiadamente forte, i.e., que não há objetos contraditórios com exceção de alguns daqueles que são produzidos por nossas práticas linguísticas e processos psicológicos. Há que se perguntar agora como formular uma lógica que incorpore (12), ou seja, que proíba contradições sobre objetos que existem independentemente de nós, mas que ainda assim seja suficientemente liberal para permitir algumas contradições. Uma alternativa seria, naturalmente, adotar lógicas distintas a depender da natureza dos objetos em questão; i.e., adotar a lógica clássica para raciocinar acerca de objetos ontologicamente independentes e uma lógica paraconsistente e paracompleta para raciocinar acerca de objetos ontologicamente dependentes. Essa proposta padece, porém, de uma séria limitação, uma vez que frequentemente lidamos com argumentos que envolvem enunciados mistos, por exemplo:

- Dom Quixote é mais famoso que Mr. Hyde.
- Mr. Hyde é mais famoso que R. L. Stevenson.
- Logo, Dom Quixote é mais famoso que R. L. Stevenson.

Para lidar com argumentos desse tipo, nos quais algumas proposições dizem respeito tanto a objetos ontologicamente dependentes quanto independentes, precisamos, portanto, de uma *única* lógica que permita raciocinar acerca de ambos os tipos de objeto indistintamente.

As *LFI*s e *LFIU*s, apresentadas na seção anterior, oferecem as ferramentas necessárias para contornar essa dificuldade, pois a presença do operador de classicalidade, $\circ$, que permite distinguir contradições explosivas de não explosivas, possibilita expressar (12) perfeitamente bem, como veremos agora.

Considere a lógica (de primeira ordem) que resulta de acrescentarmos as seguintes regras a *BLE* (cfe. seção 2.4.2):

$$\frac{}{A \vee \neg A}\, PEM \qquad \frac{\circ A \qquad A \qquad \neg A}{B}\, EXP^{\circ}$$

$$\frac{\neg\circ A}{A}\, \neg\circ E \; \frac{\neg\circ A}{\neg A}$$

$$\frac{A(c)}{\forall x A(x)}\, \forall I \qquad \frac{\forall x A(x)}{A(c)}\, \forall E \qquad \frac{A(c)}{\exists x A(x)}\, \exists I \qquad \frac{\exists x A(x) \qquad \begin{array}{c}[A(c)]\\ \vdots \\ B\end{array}}{B}\, \exists E$$

onde $A(c)$ é o resultado de se substituírem todas as ocorrências livres de x em $A(x)$ pela constante individual c e onde as regras para os quantificadores estão sujeitas às restrições usuais[14].

Essa lógica é conhecida como *LFI1* e foi introduzida pela primeira vez em (Carnielli, João Marcos, e de Amo 2000) para modelar bancos de dados contraditórios (ver também (Amo, Carnielli, e João Marcos 2002; Omori e Waragai 2011)). Note que em virtude da presença da regra *PEM*, *LFI1* não é uma *LFIU*, mas poderia ser facilmente modificada de modo que o sistema resultante fosse capaz de lidar com contextos nos quais o princípio do terceiro excluído não vale[15]. Além das regras de *BLE* e das regras para os quantificadores, *LFI1* possui duas novas regras envolvendo $\circ$, que são as duas versões de $\neg \circ E$. Elas permitem provar que:

(13) $\vdash \neg \circ A \leftrightarrow (A \wedge \neg A)$,

que expressa que sempre que se uma proposição não tem um comportamento clássico, ela é contraditória. Ao contrário do que se possa esperar, entretanto, que uma proposição A não seja contraditória não implica $\circ A$ em *LFI1*:

(14) $\nvdash \neg(A \wedge \neg A) \rightarrow \circ A$.

[14] Em $\forall I$ e $\neg\exists I$, $A(c)$ ($\neg A(c)$) não depende de nenhuma hipótese na qual c ocorra; e em $\exists E$ e $\neg\forall E$, c não ocorre em B e em nenhuma hipótese da qual $A(c)$ ($\neg A(c)$) dependa, com exceção possivelmente de $A(c)$ ($\neg A(c)$).

[15] Ver (Omori e Waragai 2011; Sano e Omori 2013; Omori e Sano 2014), onde a lógica *BS4*, que é uma versão paracompleta de *LFI1*, é investigada.

A razão é que ¬ é uma negação paraconsistente e, assim, não proíbe que $A \wedge \neg A$ e $\neg(A \wedge \neg A)$ valham simultaneamente[16].

Uma propriedade muito importante de *LFI1* é que se determinados predicados $P, Q, \ldots$ da linguagem não admitem instâncias contraditórias,

(15) $\forall x \circ P(x)$,

(16) $\forall x \forall y \circ Q(x, y)$,

então qualquer fórmula formada por meio de P e Q, e.g.,

(17) $\forall x(P(x) \rightarrow \exists y(Q(x, y) \wedge \neg P(y)))$,

também não pode ser contraditória (sob pena de trivialidade). Ou seja, sempre que (15) e (16) valem, a seguinte fórmula também vale:

(18) $\circ(\forall x(P(x) \rightarrow \exists y(Q(x, y) \wedge \neg P(y))))$.

Isso significa que em *LFI1* podemos identificar os contextos em que estamos justificados a raciocinar classicamente como aqueles nos quais todos os raciocínios envolvidos tratam de propriedades que não admitem contradições (o que, no exemplo acima, corresponde a assumir (15) e (16)), uma vez que qualquer contradição envolvendo essas propriedades levaria inevitavelmente à trivialidade, exatamente como na lógica clássica[17].

Independência ontológica e não contradição

Voltemos à questão de como expressar (12) em *LFI1*. Uma alternativa seria argumentar que é possível identificar claramente as propriedades que apenas objetos ontologicamente independentes têm e exigir que todos os pre-

[16] Na verdade, $\neg(A \wedge \neg A)$ é um teorema de *LFI1*. Assim, um par de fórmulas contraditórias, A e $\neg A$, gera automaticamente um número infinito de outras contradições: $A \wedge \neg A$ e $\neg(A \wedge \neg A)$, $((A \wedge \neg A) \wedge \neg(A \wedge \neg A))$ e $\neg(A \wedge \neg A) \wedge \neg(A \wedge \neg A))$ etc.

[17] Para um discussão filosófica mais detalhada sobre a relação entre a lógica clássica e *LFI1*, ver (Antunes 2019b).

dicados que correspondem a essas propriedades satisfaçam o seguinte esquema:

(19) $\forall x_1 \ldots \forall x_n \circ \mathbf{P}(x_1, \ldots, x_n)$,

onde **P** é um predicado arbitrário de *n* lugares. Como consequência do fato sobre *LFI1* mencionado acima, teríamos então que qualquer proposição envolvendo as propriedades em questão – e, portanto, que trate exclusivamente de objetos ontologicamente independentes – resultaria em trivialidade se contraditória. Contudo, o fato de muitas propriedades, como a de ser famoso, aplicarem-se indistintamente a objetos ontologicamente dependentes e independentes mostra que essa proposta está longe de ser ideal.

A solução consiste em restringir a aplicação do operador ∘ àquelas instâncias de **P** que são ontologicamente independentes, o que pode ser feito através do esquema:

(20) $\forall x_1 \ldots \forall x_n (I(x_1) \wedge \ldots \wedge I(x_n) \rightarrow \ \circ \mathbf{P}(x_1, \ldots, x_n))$,

onde *I* é um predicado não lógico especial que expressa à propriedade de ser ontologicamente independente.

Definindo o predicado de ser ontologicamente dependente, *D*, por meio de:

(21) $D(t) := \neg I(t)$,

(onde *t* é uma constante ou variável individual) e assumindo $\forall x \circ I(x)$ – o que evita que um mesmo objeto seja simultaneamente ontologicamente dependente e independente – podemos provar através de (13) que (20) é equivalente a:

(22) $\forall x_1 \ldots \forall x_n (\mathbf{P}(x_1, \ldots, x_n) \wedge \neg \mathbf{P}(x_1, \ldots, x_n) \rightarrow D(x_1) \vee \ldots \vee D(x_n))$.

(22) expressa que se **P** tem instâncias contraditórias, então a contradição em questão diz respeito a ao menos um objeto ontologicamente dependente.

Não é difícil provar que em *LFI1* (20) e (22) podem ser generalizados para quaisquer fórmulas da linguagem, desde que alguns cuidados envolvendo os quantificadores sejam observados. As versões generalizadas de (20) e (22) afirmam respectivamente que não há contradições não explosivas sobre objetos ontologicamente independentes e, equivalentemente, que quaisquer contradições não explosivas sempre dizem respeito a ao menos algum objeto ontologicamente dependente. Ambas são a contraparte formal de (12) em *LFI1*, pois proíbem, através da regra EXP°, quaisquer contradições envolvendo proposições sobre objetos ontologicamente independentes.

O papel da lógica clássica

Em virtude do comportamento do operador ∘ em *LFI1*, que permite recuperar a lógica clássica para as proposições que ocorrem em seu escopo, (20) (ou equivalentemente (22)) implica que qualquer inferência classicamente válida é também válida em *LFI1*, contanto que as todas proposições envolvidas tratem exclusivamente de objetos ontologicamente independentes. Por outro lado, inferências clássicas que não satisfazem essa condição podem eventualmente não serem válidas. Como consequência, ao adotar (20) (ou (22)), podemos acomodar a lógica clássica no interior de *LFI*1. Dessa perspectiva, o raciocínio clássico nada mais é do que um caso particular do paraconsistente no qual todas as entidades sobre as quais se está racionando existem independentemente de nós, de nossas práticas linguísticas e de nossos processos psicológicos. Esse fato ilustra bem o aspecto multilógico das *LFIs* e *LFUs*, que permite lidar com diferentes contextos argumentativos, que estão sujeitos a diferentes relações de consequência lógica, em um único sistema.

Mas, afinal, se não há nada que proíba descrições contraditórias de objetos ontologicamente dependentes, por que então estamos, em geral, inclinados a encarar contradições com extrema suspeição, independentemente da natureza dos objetos sobre os quais elas dizem respeito? A resposta a essa pergunta tem a ver com a maneira como interagimos com objetos ontologicamen-

te dependentes, pois na maior parte das vezes eles são produzidos de modo a refletir as características de objetos ontologicamente independentes. Por exemplo, uma personagem de ficção é em geral criada para representar de modo mais ou menos fidedigno as características de uma pessoa (ou animal) real. Assim, é natural, embora não seja necessário, que a personagem em questão acabe não sendo descrita de maneira contraditória. Esse fato explica também por que estamos frequentemente inclinados a pensar que inferências clássicas valem irrestritamente, independentemente do domínio sobre os quais estamos raciocinando – pois, como acabamos de ver, todas as inferências clássicas são válidas sempre que as proposições envolvidas tratam exclusivamente de objetos ontologicamente independentes.

Em suma: via de regra, interagimos com objetos ontologicamente dependentes *como se* eles fossem ontologicamente independentes, o que explica por que na maior parte das vezes resistimos erroneamente à ideia de que eles possam ter propriedades contraditórias e por que tomamos, também erroneamente, as inferências da lógica clássica como sendo irrestritamente válidas.

Considerações finais

Vimos aqui duas abordagens não dialeteístas à paraconsistência. Na primeira, que propõe uma interpretação epistêmica das lógicas paraconsistentes, uma contradição envolvendo A e $\neg A$ indica a presença de evidências conflitantes, porém não conclusivas, para A e $\neg A$. A ocorrência de contradições em um dado contexto argumentativo não significa, portanto, que contradições sejam verdadeiras, como propõem os dialeteístas, uma vez que a noção de evidência é mais fraca do que a de verdade. Na segunda abordagem, que possui inspiração em posições nominalistas em filosofia da matemática, contradições são autorizadas contanto que elas tratem de objetos ontologicamente dependentes, que, grosso modo, são aqueles produzidos por nossos processos psicológicos e práticas linguísticas.

Em ambas as abordagens, não há um conflito real entre as lógicas clássica e paraconsistente. Enquanto na interpretação epistêmica as diferenças entre as duas lógicas se devem ao fato de elas tratarem de noções de

consequência lógica distintas (viz., preservação de verdade vs. preservação de evidência), na interpretação nominalista a lógica clássica é vista como um caso particular da lógica paraconsistente no qual todos os objetos envolvidos existem independentemente de nós. Além disso, a interpretação epistêmica combinada com o operador de classicalidade ◦ permite recuperar todas as inferências clássicas para as proposições conclusivas.

Contradições produzidas por objetos ontologicamente dependentes podem, em certo sentido, ser consideradas verdadeiras. Mas nesse caso, não se trata de uma noção de verdade estritamente realista. Desse modo, alguém poderia, por exemplo, dizer que as proposições "as personagens de Jules e Jim são a mesma" e "as personagens de Jules e Jim não são a mesma" são ambas verdadeiras, dado que "Jules e Jim são a mesma pessoa" e "Jules e Jim não são a mesma pessoa" são verdadeiras na pequena história contada por (Everett 2005). Poder-se-ia, então, distinguir um dialeteísmo forte, relacionado a contradições ontologicamente independentes, de um dialeteísmo fraco, devido a contradições ontologicamente dependentes. Nós, entretanto, preferimos considerar que a atribuição de verdade a proposições sobre objetos ficcionais (e também a proposições sobre objetos ontologicamente dependentes) são antes apenas um 'modo de falar', do que uma posição legitimamente dialeteísta.

Embora seja uma tese controversa, o dialeteísmo representa uma importante posição em filosofia da lógica. Entretanto, o dialeteísmo não é a única e, de acordo com os autores, nem a melhor maneira de compreender a ocorrência de contradições em determinados contextos argumentativos. Como vimos aqui, existem diferentes formas de interpretá-las sem assumir que o mundo ou a realidade tenha aspectos intrinsecamente contraditórios de modo tornar algumas contradições verdadeiras.

Referências

de Amo, S., W. Carnielli, João Marcos. 2002. "A Logical Framework for Integrating Inconsistent Information in Multiple Databases." In *Foundations of Information and Knowledge Systems*, editado por Thomas Eiter e Klaus-Dieter Schewe, 67–84. Springer.

Antunes, H. 2018. "On Existence, Inconsistency, and Indispensability." *Principia: An International Journal of Epistemology* 22: 7–34.

———. 2019a. "Contradictions for Free: Towards a Nominalistic Interpretation of Contradictory Theories." Tese de Doutorado, Universidade de Campinas. https://bit.ly/antunes_phdthesis.

———. 2019b. "Enthymematic Classical Recapture." *The Logic Journal of IGPL*. https://doi.org/0.1093/jigpal/jzy061.

Antunes, H., W. Carnielli, A. Kapsner, A. Rodrigues. Kripke-style modelsfor logics of evidence and truth. Axioms, 9:1–16, 2020.

Aristotle. 1941. "Translation of Aristotle's Metaphysics." In *The Basic Works of Aristotle*, editado por R. McKeon, traduzido por W. D. Ross. New York: Random House.

Azzouni, J. 2004. *Deflating Existential Consequence: A Case for Nominalism*. Oxford University Press.

———. 2015. "Mathematical Fictions." In *Fiction and Art: Explorations in Contemporary Theory*, editado por Anantha Ch. Sukla, 63–77. Bloomsbury.

Batens, D. 2002. "In Defence of a Programme for Handling Inconsistencies." In *Inconsistency in Science*. Springer.

Belnap, N. D. 1977a. "A Useful Four-Valued Logic." In *Modern Uses of Multiple Valued Logics*, editado por G. Epstein e J. M. Dunn. D. Reidel, Dordrecht.

———. 1977b. "How a Computer Should Think." In *Contemporary Aspects of Philosophy*. Oriel Press.

Carnielli, W., M. E. Coniglio. 2016. *Paraconsistent Logic: Consistency, Contradiction and Negation*. Vol. 40 of Logic, Epistemology, and the Unity of Science series. Springer.

Carnielli, W., M. E. Coniglio, João Marcos. 2007. "Logics of Formal Inconsistency." In *Handbook of Philosophical Logic (2nd. Edition)*, editado por D. M. Gabbay e F. Guenthner, 14:1–93. Springer.

Carnielli, W., M. E. Coniglio, A. Rodrigues. 2019. "Recovery Operators, Paraconsistency and Duality." *Logic Journal of the IGPL*. https://doi.org/10.1093/jigpal/jzy054.

Carnielli, W., João Marcos, S. de Amo. 2000. "Formal Inconsistency and Evolutionary Databases." *Logic and Logical Philosophy* 8: 115–52.

Carnielli, W., A. Rodrigues. 2016. "On the Philosophy and Mathematics of the Logics of Formal Inconsistency." In *New Directions in Paraconsistent Logic*, editado por J.-Y. Beziau, M. Chakraborty, e S. Dutta, 57–88. Springer.

———. 2017. "An Epistemic Approach to Paraconsistency: A Logic of Evidence and Truth." *Synthese* 196: 3789–3813. https://doi.org/10.1007/s11229-017-1621-7.

———. 2020. "On Epistemic and Ontological Interpretations of Intuitionistic and Paraconsistent Paradigms." *Logic Journal of the IGPL*. https://doi.org/10.1093/jigpal/jzz041.

Chateaubriand, O. 2001. *Logical Forms Vol. 1*. Campinas: UNICAMP-CLE.

Colyvan, M. 2008. "The Ontological Commitments of Inconsistent Theories." *Philosophical Studies* 141: 115–23.

da Costa, N. C. A. 1963. "Sistemas Formais Inconsistentes (Inconsistent Formal Systems, in Portuguese)." Tese de Habilitação. Universidade Federal do Paraná, Curitiba, Brazil.

———. 1974. "On the Theory of Inconsistent Formal Systems." *Notre Dame Journal of Formal Logic* XV (4): 497–510.

da Costa, N. C. A., e S. French. 2003. *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press.

da Costa, N. C. A., D. Krause. 2014. "Physics, Inconsistency, and Quasi-Truth." *Synthese* 191 (13): 3041–55.

Costa, N. C. A. da, D. Krause e O. Bueno. 2007. "Paraconsistent Logics and Paraconsistency." In *Philosophy of Logic – Handbook of the Philosophy of Science Vol. 5*, editado por D. Jacquette et al. Elsevier.

Dunn, J. M. 1976. "Intuitive Semantics for First-Degree Entailments and 'Coupled Trees'." *Philosophical Studies* 29: 149–68.

Everett, A. 2005. "Against Fictional Realism." *The Journal of Philosophy* 102: 624–49.

Frade, L., A. Rodrigues. 2020. "Some Remarks on Logical Realism and Logical Pluralism." https://bit.ly/realism_pluralism.

Frápolli, M. 1996. "Review of Graham Priest 'Beyond the Limits of Thought'." *The Review of Modern Logic*, no. 6: 437–39.

Frege, G. 1879. "Begriffsschrift, a Formula Language, Modeled Upon That of Arithmetic, for Pure Thought." In *From Frege to gödel*, editado por J. van Heijnoort, 1–82. Lincoln: toExcel Press (1999).

Garfield, J. 2004. "To Pee and Not to Pee? Could *That* Be the Question? (Further Reflections of the Dog)." In *The Law of Non-Contradiction: New Philosophical Essays*, editado por G. Priest, J. C. Beall, e B. Armour-Garb. Oxford University Press.

Gottlieb, P. 2013. "Aristotle on Non-Contradiction." *The Stanford Encyclopedia of Philosophy*.

Kiefer, N. M., J. M. Dunn. 2019. "Contradictory Information: Better Than Nothing? The Paradox of the Two Firefighters." In *Graham Priest on Dialetheism and Paraconsistency*, editado por C. Başkent e T. M. Ferguson. Springer.

Łukasiewicz, J. 1910. "On the Principle of Contradiction in Aristotle." *The Review of Metaphysics* Vol. 24, No. 3 (1971): 485–509.

Mares, E. 2004. "Semantic Dialetheism." In *The Law of Non-Contradiction: New Philosophical Essays*, editado por Beall Priest e Armour-Garb. Oxford University Press.

Meheus, J., ed. 2002. *Inconsistency in Science*. Dordrecht: Springer.

Nickles, T. 2002. "From Copernicus to Ptolemy: Inconsistency and Method." In *Inconsistency in Science*, editado por J. Meheus. Dordrecht: Springer.

Omori, H., K. Sano. 2014. "da Costa Meets Belnap and Nelson." In *Recent Trends in Philosophical Logic*, editado por R. Ciuni, H. Wansing, e C. Willkommen. Springer.

Omori, H., T. Waragai. 2011. "Some Observations on the Systems Lfi1 and Lfi1*." In *Proceedings of Dexa2011*, editado por F. Morvan, A. M. Tjoa, and R. Wagner, 320–24. IEEE Computer Society.

Popper, K. 1963. *Conjectures and Refutations*. New York, Harper.

Priest, G. 1998. "What Is so Bad about Contradictions?" *Journal of Philosophy* 95: 410–26.

———. 2006. *In Contradiction: A Study of the Transconsistent (2nd Edition)*. Oxford University Press.

———. 2019. "Some Comments and Replies." In *Graham Priest on Dialetheism and Paraconsistency*.

Priest, G., F. Berto, W. Zach. 2018. "Dialetheism." *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*.

Rodrigues, A. 2017. "Uma Breve História Das Contradi: De Aristóteles à Paraconsistência." *Nuntius Antiquus* 13: 241–62.

Rodrigues, A., J. Bueno-Soler, W. Carnielli. 2020. "Measuring Evidence: A Probabilistic Approach to an Extension of Belnap-Dunn Logic." *Synthese*. https://doi.org/10.1007/s11229-020-02571-w.

Sano, K., H. Omori. 2013. "An Expansion of First-Order Dunn-Belnap Logic." *Logic Journal of the IGPL* 22: 458–81.

Steinberger, F. 2016. "Explosion and the Normativity of Logic." *Mind* 125: 385–419.

Szmuc, D. 2016. "Defining LFIs and LFUs in Extensions of Infectious Logics." *Journal of Applied Non-Classical Logics* 26 (4): 286–314.

Tarski, A. 1956. "The Concept of Truth in Formalized Languages (1936)." In *Logic, Semantics, Metamathematics*. Clarendon Press.

Tugendhat, E., U. Wolf. 1996. *Propedêutica lógico-Semântica*. Petrópolis: Editora Vozes.

Wansing, H., S. Odintsov. 2016. "On the Methodology of Paraconsistent Logic." In *Logical Studies of Paraconsistent Reasoning in Science and Mathematics*, editado por H. Andreas e P. Verdée. Springer.

O PARADOXO DO MENTIROSO: ASPECTOS TÉCNICOS E FILOSÓFICOS

Guilherme Cardoso[1]

Ederson Melo[2]

Túlio Aguiar[3]

RESUMO

Nosso objetivo principal, neste texto, é o de fornecer uma introdução crítica ao paradoxo do Mentiroso, tanto do ponto de vista técnico, quanto do ponto de vista filosófico. Primeiramente, vamos apresentar diferentes versões do Mentiroso e relacionar importantes resultados limitativos associados a ele, o Lema Diagonal e o Teorema da Indefinibilidade de Tarski. Em seguida, iremos apresentar algumas das principais abordagens do paradoxo. Finalmente, iremos confrontá-las com uma mesma objeção: a aparente impossibilidade de se conciliar completamente poder expressivo ilimitado de um lado e consistência ou não-trivialidade de outro.

Palavras-chave: Paradoxo do Mentiroso. Lema Diagonal. Teorema da Indefinibilidade. Lacunas. Aglutinações. Trivialidade.

ABSTRACT

Our main goal in this text is to provide a critical introduction to the Liar paradox, both from a technical point of view, as well as from a philosophical point of view. First, we shall present different versions of the Liar and list important limiting results associated with it, the Diagonal Lemma and Tarski's Undefinability Theorem. Next, we shall present some of the main approaches to the paradox. Finally, we shall confront them with the same challenge: the apparent impossibility of completely reconciling unlimited expressive power on the one hand and consistency or non-triviality on the other.

Keywords: Liar Paradox. Diagonal Lemma. Undefinability Theorem. Gaps. Gluts. Triviality.

[1] Professor do Departamento de Filosofia da Universidade Federal de Ouro Preto.
E-mail: guilherme.cardoso@ufop.edu.br. ORCID:.0000-0003-0045-8508.
[2] Professor do Departamento de Filosofia da Universidade Federal do Maranhão.
E-mail: edersonsafra@yahoo.com.br. ORCID: 0000-0003-1706-1128.
[3] Professor do Departamento de Filosofia da Universidade Federal de Minas Gerais.
E-mail: taguiar.bh67@gmail.com. ORCID: 0000-0003-1991-5406.

1. Introdução

Para começar, considere o seguinte problema: imagine que você acaba de chegar em uma ilha onde todos os nativos são cavaleiros ou escudeiros. Os cavaleiros são inteiramente honestos e os escudeiros são inteiramente mentirosos; ou seja, tudo que os cavaleiros dizem é verdadeiro e tudo que os escudeiros dizem é falso. Você está à procura de John, um escudeiro da ilha que lhe deve dinheiro. A certa altura, você suspeita que o tenha encontrado, mas não tem certeza, pois John tem dois irmãos gêmeos, James e William. James também é um escudeiro e William é um cavaleiro. Suponha que nessa situação você tenha direito a uma única pergunta com menos de cinco palavras para descobrir se o indivíduo é o John. Qual é a pergunta? A partir de uma breve reflexão, podemos ver que a resposta para a pergunta em questão deveria ser: "você é o James?". Somente John responderá que sim. James e William, forçosamente, deverão dizer que não. Esse problema foi proposto por Smullyan (1978), como um enigma sobre a verdade. Agora, suponha que John (ou qualquer outro habitante da ilha) afirmasse o seguinte:

(*l*) Eu sou um escudeiro.

Ora, pelo que sabemos a respeito da ilha, John deve ser um cavaleiro ou um escudeiro. Ele é um cavaleiro se, e somente se,[4] tudo o que ele diz é verdadeiro, ele é um escudeiro sse tudo o que ele diz é falso. John afirma que é um escudeiro. Por um lado, se a sentença (*l*) for verdadeira, então John é um escudeiro (é exatamente isso que diz a sentença) e, assim sendo, (*l*) deve ser falsa (já que ela foi dita por um escudeiro que somente profere falsidades). Por outro lado, se a sentença (*l*) for falsa, então John não é um escudeiro e, como só há escudeiros ou cavaleiros na ilha, John deve ser, então, um cavaleiro e, sendo assim, a sentença deve ser verdadeira (já que ela foi dita por um escudeiro que somente profere falsidades). Portanto, John é um cavaleiro sse ele é um escudeiro, sua afirmação é verdadeira sse sua afirmação é falsa.

Diferentemente do enigma apresentado na abertura do presente texto, o problema de determinar o valor de verdade de (*l*) não parece ter solução nas

[4] Deste ponto em diante, vamos abreviar a expressão "se, e somente se" por "sse".

condições dadas. Nesse sentido, dizemos que o segundo problema é um paradoxo. Especificamente, esse paradoxo pode ser considerado uma versão do paradoxo do Mentiroso, foco do presente artigo.

Um *paradoxo* pode ser entendido como um argumento aparentemente aceitável que conduz, através de premissas (ou pressuposições) aparentemente aceitáveis, a uma conclusão aparentemente inaceitável. Nessa perspectiva, dizemos que um paradoxo surge quando somos levados para o aparentemente inaceitável partindo do aparentemente aceitável (cf. READ, 1995, p.193 e SAINSBURY, 2009, p.1).

Uma solução rápida e aceitável do paradoxo do escudeiro consiste em simplesmente recusar sua existência. Uma ilha que satisfaça rigorosamente tais condições simplesmente não existe, ela é tão plausível quanto o cavalo branco que não é branco. Como veremos em seguida, uma família de paradoxos pode ser associada ao Escudeiro, a família do paradoxo do Mentiroso.

Como veremos neste artigo, há várias versões do paradoxo do Mentiroso que não permitem uma solução tão simples. Particularmente, várias abordagens dadas ao paradoxo do Mentiroso surgem ao propor algum tipo de revisão dos pressupostos envolvidos. Ainda que possamos tratar os paradoxos de maneira lúdica, como no exemplo do Escudeiro, alguns paradoxos podem revelar problemas fundamentais sobre os conceitos envolvidos.

Nesta área da Lógica e da Filosofia, tão importantes quanto os próprios paradoxos são as ricas abordagens e técnicas elaboradas no intuito de resolvê-los. Além disso, importantes técnicas semânticas e importantes resultados metalógicos foram estabelecidos como consequência dos limites estabelecidos por estes paradoxos em teorias formais.

Tendo isso em vista, o presente artigo tem como objetivo apresentar uma introdução crítica ao paradoxo do Mentiroso (ou, simplesmente, Mentiroso), destacando os principais resultados técnicos a ele relacionados, as principais abordagens propostas ao paradoxo e levantando algumas reflexões filosóficas relacionadas.

Na seção 2, vamos expor os pressupostos envolvidos no Mentiroso a partir de suas principais versões, reunidas no conjunto de paradoxos que cha-

mamos aqui de “família do mentiroso”. Na seção 3, veremos que as versões do Mentiroso apresentadas na seção 2 estão apoiadas em resultados formais importantes, geralmente estudados em cursos de metalógica. Especificamente, iremos conferir o Lema da Diagonal e o Teorema da Indefinibilidade de Tarski. Feito isso, na seção 4, iremos apresentar algumas das principais abordagens filosóficas do Mentiroso. Particularmente, vamos expor três influentes tratamentos dados a ele. Começaremos pela abordagem de Tarski que lida com o Mentiroso através de uma hierarquias de linguagens. Posteriormente, veremos a abordagem de Kripke que admite lacunas (*gaps*) de valores de verdade (i.e. sentenças que não são nem verdadeiras e nem falsas). Por fim, vamos conferir a abordagem dialeteísta dada por Priest que assume aglutinações (*gluts*) de valores de verdade (i.e. sentenças que são tanto verdadeiras quanto falsas). Ao longo da seção 4 vamos indicar que as abordagens dadas ao Mentiroso, inclusive a dialeteísta, restringem o poder expressivo da linguagem de uma maneira ou outra e que, portanto, a resiliência do Mentiroso está relacionada com o limite expressivo das linguagens em que ele ocorre.

2. A Família do Mentiroso

Nesta seção, iremos apresentar informalmente as principais versões do Mentiroso, apontando seus pressupostos.

A versão simples do Mentiroso, também conhecida como Mentiroso Simples, pode ser exemplificada por qualquer sentença autorreferencial que atribui a si mesma o predicado "é falsa". Considere, por exemplo, a seguinte sentença:

(1) (1) é falsa.

(1) é uma sentença que diz de si mesma que é falsa.[5] Se (1) é verdadeira, o que ela diz é o caso, então (1) é falsa. Se, por outro lado, (1) é falsa, o que ela diz é o caso, então ela é verdadeira. (1) é verdadeira sse (1) é falsa. Logo, (1) é verdadeira e falsa. Dois pressupostos fundamentais acerca do conceito de

[5] Existem outros mecanismos disponíveis na linguagem natural para produzir sentenças autorreferenciais. Por exemplo, poderíamos dizer que a sentença em itálico na nota 5 do presente texto é falsa.

verdade estão implícitos no presente argumento, o Princípio de Bivalência (PB) e a Transparência da Verdade (ou Esquema T):

(PB) Toda sentença é verdadeira ou falsa.

(T) Uma sentença é verdadeira sse o que ela diz é o caso.

À primeira vista, pode-se supor que a rejeição do (PB) seja completamente suficiente para resolver o Mentiroso. Esta é uma suposição apressada, como fica claro pela apresentação de uma versão mais forte do Mentiroso, o Mentiroso Reforçado:

(2) (2) não é verdadeira.

Na medida em que se rejeita o (PB), algumas sentenças não são verdadeiras nem falsas. Classicamente, "falsa" e "não-verdadeira" são predicados co-extensionais. Esse tipo de rejeição do (PB) é normalmente identificado como *truth value gaps* (lacunas de valores-verdade). Entretanto, apenas classificar (2) como nem verdadeira nem falsa não resolve o problema. Afinal, se (2) não é verdadeira nem falsa, *a fortiori* (2) não é verdadeira, confirmando aquilo que é dito por (2). O novo paradoxo pressupõe não o (PB), mas o Princípio do Terceiro Excluído (PTE); que de toda sentença, ela ou sua negação é verdadeira. Novamente, por (T), (2) é verdadeira sse (2) não é verdadeira. Logo, pelo (PTE), (2) é verdadeira e (2) não é verdadeira.

Nas versões anteriores, os paradoxos foram obtidos por meio de sentenças que referiam-se a si mesmas, sentenças autorreferenciais. Entretanto, algumas versões do Mentiroso podem ser também engendradas por ciclos de referência, como exemplificado pelo paradoxo do Cartão Postal:

(3) (4) é verdadeira.

(4) (3) não é verdadeira.

Note que neste caso não temos autorreferência, mas ainda temos algum tipo de circularidade sob a relação de referir, (3) se refere a (4) que, por sua vez, refere-se a (3). Por argumentos análogos aos anteriores, podemos verificar que (3) e (4) também não admitem avaliações consistentes.

A circularidade presente em (1), (2), (3) e (4) parece depender de certas circunstâncias empíricas e extrínsecas em relação às sentenças elas mesmas. Entretanto, como veremos mais tarde (por meio do Lema Diagonal), é possível garantir a existência desse tipo de circularidade mesmo em cenários muito bem arregimentados.

A próxima versão, conhecida como paradoxo de Yablo,[6] aparenta dispensar todo tipo de circularidade ou autorreferência. Neste caso, vamos precisar de uma lista com infinitas sentenças:

(5_1) Para todo $x>1$, (5_x) não é verdadeira.

(5_2) Para todo $x>2$, (5_x) não é verdadeira.

(5_3) Para todo $x>3$, (5_x) não é verdadeira.

$\vdots$

(5_n) Para todo $x>n$, (5_x) não é verdadeira.

(5_{n+1}) Para todo $x>n+1$, (5_x) não é verdadeira.

$\vdots$

Suponha que (5_k) é verdadeira (para qualquer $k>0$). Assim, para todo $x>k$, (5_x) não é verdadeira, pois é isto mesmo que (5_k) diz. Consequentemente, (5_{k+1}) não é verdadeira. Ora, se (5_{k+1}) não é verdadeira, então, para algum $y>k+1$, (5_y) é verdadeira, logo, (5_k) não é verdadeira. Portanto, nenhuma sentença da lista pode ser consistentemente avaliada como verdadeira, nenhuma delas é verdadeira. Isto implica que, para qualquer $x>k$, (5_x) não é verdadeira, qualquer que seja $k>0$. Logo, (5_k) é verdadeira, qualquer que seja $k>0$.

[6] Inicialmente proposta por Yablo (1993).

Em suma, todas as sentenças da lista são verdadeiras e nenhuma delas é verdadeira.

O paradoxo de Yablo sugere que a circularidade pode não ser um requisito necessário para a obtenção dos paradoxos.[7] Veremos agora que a circularidade também não é, alegadamente, um requisito suficiente para a obtenção dos paradoxos. Algumas sentenças podem engendrar circularidade sob a relação de referir, sem contudo derivar paradoxos. Considere as seguintes sentenças:

(6) (6) é um exemplo de autorreferência.

(7) (7) contém menos de vinte palavras.

(8) (8) é verdadeira.

As três últimas sentenças não parecem engendrar paradoxos. Ao menos, não parece haver razões que nos impedissem de avaliá-las consistentemente como verdadeiras ou falsas. (6) e (7) são intuitivamente verdadeiras. Quanto a (8), não há qualquer critério para decidir se ela é verdadeira ou falsa. De todo modo, ela pode ser avaliada tanto como verdadeira quanto como falsa, sem que isto implique uma contradição. Na literatura, (8) é conhecida como *Truth Teller*. O valor de (8) é frequentemente classificado como *up for grabs* ou, simplesmente, como um valor em disputa, ou arbitrário.

Algumas versões do Mentiroso podem depender de fatores contingentes e extrínsecos às sentenças elas mesmas. Assim, existem sentenças que envolvem essencialmente circularidade, mas que podem ou não engendrar paradoxos, a depender de fatores extrínsecos. Estas versões são conhecidas como Mentirosos Contingentes. Exemplo:

(9) (9) é falsa e não existe vida em Trappist-1.

Ora, se (9) é verdadeira, então, as duas sentenças que a compõem devem ser verdadeiras. Assim, (9) não pode ser verdadeira (é falsa). Por outro lado, se ao

[7] Isso não é consensual. Pode-se argumentar que cada uma das sentenças na lista infinita de Yablo é autorreferencial. Grosso modo, se cada uma das sentenças diz de todas as seguintes que elas não são verdadeiras, então cada uma delas tem um componente indexical que só pode ser preenchido por referência à sentença ela mesma. Cada sentença diz algo como "todas as sentenças seguintes a esta própria sentença são não-verdadeiras".

menos uma de suas componentes é falsa, (9) é falsa. Suponha que existe vida em Trappist-1. Neste caso, (9) é simplesmente falsa. Isto torna verdadeira a primeira componente de (9), mas não torna verdadeira (9), ela mesma. Alternativamente, se não existe vida em Trappist-1, (9) é verdadeira sse (9) é falsa. Portanto, o valor semântico de (9) depende de fatos contingentes e extrínsecos à sentença ela mesma. Se houver vida em Trappist-1, (9) é falsa, de outro modo, (9) é paradoxal.

Assim como podemos utilizar a conjunção e outros conectivos ou quantificadores para formar novos mentirosos contingentes, podemos também utilizar os conectivos e quantificadores para construir versões do paradoxo de Curry. A especificidade do Curry é que ele coloca diretamente uma ameaça de trivialidade para as teorias.[8] Considere o seguinte exemplo:

(10) Se (10) é verdadeira, então todas as sentenças são verdadeiras.

Pela interpretação clássica da condicional, (10) não pode ser avaliada consistentemente como falsa, pois se (10) é falsa, sua antecedente é verdadeira e sua consequente é falsa e, assim, (10) é verdadeira e falsa. Por outro lado, se (10) é verdadeira, então, a sentença antecedente de (10) também é verdadeira, pois é isso mesmo que ela diz. Por *Modus Ponnens* (MP), todas as sentenças são verdadeiras. Vimos que, sob a hipótese de que (10) é verdadeira, então todas sentenças são verdadeiras. Ora, isto é exatamente o que (10) diz. Isso é uma demonstração da verdade de (10).[9] Logo, por (MP), todas as sentenças são verdadeiras. Note que poderíamos substituir a consequente de (10) por qualquer sentença da linguagem em questão e o mesmo argumento poderia ser utilizado como uma demonstração desta sentença.

O paradoxo de Curry desempenha, em relação à estratégia de aglutinações, um papel análogo ao que o Mentiroso Reforçado desempenha em relação à estratégia de lacunas. A estratégia de aglutinações consiste em admitir

[8] Classicamente, inconsistência implica trivialidade, pelo Princípio de Explosão (PE). O paradoxo de Curry ameaça de trivialidade mesmo aquelas teorias que rejeitam o (PE), as teorias paraconsistentes.

[9] Formalmente, esse passo é desempenhado pelo Teorema da Dedução (TD) (ou Prova Condicional). Em alguns contextos, o mesmo argumento pode ser formalizado por meio da Contração (Con).

que algumas sentenças são concomitantemente verdadeiras e falsas (*truth-value gluts*). O paradoxo de Curry mostra que, nestes casos, se ainda valem certas propriedades da condicional, como (MP) e (Con) ou (TD), a trivialidade não pode ser evitada.

Existem muitas outras versões do Mentiroso, mas seria impossível introduzir de forma exaustiva toda esta família de paradoxos. Tendo apresentado informalmente as principais versões, tentaremos agora precisar um pouco melhor os principais resultados formais que fundamentam estes argumentos.

3. Aspectos formais

Os paradoxos acima apresentados estão apoiados em certos resultados formais estudados em cursos de Metalógica.[10] Nesta seção, iremos apresentar alguns destes resultados. Iremos apresentar o Lema Diagonal e o Teorema da Indefinibilidade de Tarski. O Lema Diagonal garante a possibilidade de reconstruir, em teorias formais, sentenças autorreferenciais (ou que se comportam logicamente como sentenças autorreferenciais). O Teorema da Indefinibilidade estabelece um limite rigoroso para tais teorias. Grosso modo, uma teoria formal que satisfaça as condições do Lema Diagonal não pode conter o seu próprio predicado-verdade, nenhum dos seus predicados define o conjunto de todas as sentenças verdadeiras da própria teoria. Este importante resultado foi estabelecido pelo filósofo, lógico e matemático polonês Alfred Tarski.

Uma teoria é um conjunto de sentenças Γ fechado sob a relação de consequência lógica sintática, ou seja, $A \in \Gamma$, se $\Gamma \vdash A$ (para qualquer sentença A da linguagem da teoria). Note que, dado qualquer conjunto de sentenças Γ de uma linguagem L, o conjunto das suas consequências dedutivas (teoremas), $TH(\Gamma)$, é uma teoria nesta linguagem. Dizemos que uma teoria Σ (na linguagem L) é recursivamente axiomatizável, se existe um conjunto recursivo de sentenças Γ de L, tal que, $TH(\Gamma) = \Sigma$.[11]

[10] O leitor pode encontrar uma exposição mais detalhada destes resultados em Boolos at al (2007) e também em Cardoso (2018).

[11] Pela Tese Church-Turing, as funções recursivas (equivalentemente, as funções computáveis por máquinas de turing) são exatamente as funções efetivamente computáveis, aquelas que podem ser computadas por algoritmos. A função característica de um conjunto determi-

Uma teoria Σ é inconsistente sse existe uma sentença A da linguagem de Σ, tal que, $\Sigma \vdash A \wedge \neg A$. Σ é consistente sse Σ não é inconsistente. Uma teoria Σ é trivial sse, para toda sentença B da linguagem de Σ, $\Sigma \vdash B$. Pelo (PE), se Σ é inconsistente, Σ é trivial. Por outro lado, se Σ é trivial, então, Σ prova todas as sentenças da sua linguagem. Logo, *a fortiori*, Σ é inconsistente. Logo, Σ é inconsistente sse Σ é trivial.

3.1. O Lema Diagonal

Os resultados limitativos que iremos apresentar a seguir dizem respeito a um tipo especial de teorias, aquelas que são capazes de representar todas as funções recursivas. Grosso modo, isso permite que tais teorias sejam capazes de emular as sentenças paradoxais exemplificadas na seção 2. Como exemplos paradigmáticos, veremos que algumas teorias aritméticas (todas as extensões da Aritmética de Robinson - Q) satisfazem este requisito.[12]

As sentenças paradoxais exemplificadas na seção anterior não são sentenças aritméticas. A princípio, sentenças aritméticas envolvem termos para números e operações como soma, multiplicação, etc. Como podemos, então, dizer que é possível de algum modo emular as sentenças paradoxais mencionadas? Ora, sabidamente, o poder expressivo da Aritmética reside em seu potencial representativo. Por meio de codificações, podemos representar muitas outras coisas com as sentenças aritméticas. Em particular, todas as funções recursivas são representáveis em todas as extensões de Q.

A linguagem da Aritmética é definida pelo seguinte conjunto de símbolos não lógicos: $\{0, s, +, \times, <\}$. A interpretação (modelo) padrão dessa linguagem toma como domínio o conjunto dos números naturais, interpreta s como a função de sucessor, + e × como as operações de soma e multiplicação e < como a relação de ordem estrita "menor que". O conjunto de todas as sen-

na, para todo *x*, se *x* pertence ou não a este conjunto. O conjunto é recursivo sse a sua função característica é recursiva. Isso é uma simplificação. Para mais detalhes, ver Boolos et al (2007).

[12] Q é também conhecida como Aritmética Minimal. Ela é forte o suficiente para representar todas as funções recursivas, mas fraca o suficiente para não autorizar as generalizações obtidas por provas indutivas (Q não tem um axioma de indução). Esta é a principal diferença entre Q e sua mais famosa extensão, a Aritmética de Peano - PA. O leitor pode encontrar uma apresentação de Q e também de PA em Boolos et al (2007, cap.16).

tenças da linguagem da Aritmética que são verdadeiras no modelo padrão é a teoria conhecida como Aritmética Verdadeira (V). Um dos resultados que veremos a seguir diz exatamente que V não é recursivamente axiomatizável.

Antes de falar em representabilidade, precisamos adotar algum esquema de codificação. A seguinte correspondência c entre números e símbolos primitivos na linguagem da Aritmética é o primeiro passo. Vamos nos restringir a representar funções sintáticas definidas sobre expressões da própria linguagem da Aritmética. De outro modo, teríamos de incluir mais coisas nesse esquema de codificação:

(	)	,	¬	∨	∃	=	x_i	<	0	s	+	×
1	3	5	7	9	11	13	2.5^i	36	8	24	72	360

O esquema acima, portanto, permite relacionar cada símbolo primitivo da linguagem da Aritmética com um único número e cada sequência de símbolos com uma única sequência de números. Isso é suficiente para obter uma função que leve de sequências de símbolos a números, a função número de gödel:

$$gn((s_0, s_1, ..., s_{n-1})) = 2^n \times \left(\prod_{i=1}^{n-1} (\pi(i)^{c(s_{i-1})}) \right)$$

Onde π é uma função que leva de números inteiros a números primos, uma enumeração dos primos. A função c, também utilizada na definição, é aquele esquema de correspondência descrito na tabela acima. Assim, $c(s_{i-1})$ nos dá o número que corresponde ao símbolo que aparece na posição $i-1$ da sequência de símbolos. Este método de codificação está baseado no fato de que todo número natural tem uma única decomposição em fatores primos. Cada fórmula nos dá uma sequência de símbolos. Primeiramente, detectamos o cumprimento da fórmula (quantidade de símbolos) na posição de expoente do número 2. Uma sequência do tipo $(s_0, s_1, ..., s_{n-1})$ tem n termos. Depois, os símbolos são listados nas respectivas posições de expoentes dos primos. Assim, por exemplo, a fórmula $\neg \exists x_0 (x_0 < 0)$ será codificada pelo número:

$$2^8 \times 3^7 \times 5^{11} \times 7^2 \times 11^1 \times 13^2 \times 17^{36} \times 19^8 \times 23^3$$

Este é o seu número de gödel. Note que números são objetos do domínio de modelos da linguagem da Aritmética. O modo de se referir indiretamente, portanto, aos objetos codificados por números de gödel é através dos respectivos numerais, estes sim termos da linguagem da Aritmética. Vamos distinguir números de numerais por meio da notação em negrito. Assim, **0** é o numeral que denota, na interpretação padrão, o número zero.[13] Ademais, $\ulcorner e \urcorner$ é o numeral do número de gödel da expressão e.

Todavia, para que sejamos capazes de falar indiretamente sobre fórmulas na linguagem da Aritmética é antes necessário que possamos representar as funções e relações definidas sobre elas. Por exemplo, para dizer que $\neg(s(0)<0)$ é o resultado de se substituir todas as ocorrências de x_0 por $s(0)$ na fórmula $\neg(x_0<0)$, temos também de representar a relação entre estas expressões, não basta codificá-las. Isto nos leva diretamente à definição seguinte.

> **Representabilidade:** seja f uma função de aridade n e Σ uma teoria aritmética de primeira ordem. Assim, f é representável em Σ sse existe uma fórmula Φ (de aridade $n+1$) na linguagem da Aritmética, tal que,
>
> $f(k_1,\ldots,k_n)=m$ sse $\Sigma \vdash \forall y(\Phi(\ulcorner k_1 \urcorner,\ldots,\ulcorner k_n \urcorner, y) \leftrightarrow y=\ulcorner m \urcorner)$[14]

Definibilidade é a contraparte semântica da noção de representabilidade. No caso das teorias aritméticas, diremos que a função f é aritmeticamente definível sse f é representável em V. Ou seja, $f(k_1,\ldots,k_n)=m$ sse $\Phi(\ulcorner k_1 \urcorner,\ldots,\ulcorner k_n \urcorner,\ulcorner m \urcorner)$ é verdadeira no modelo padrão da linguagem da Aritmética.

Como dissemos antes, todas as funções recursivas são representáveis nas extensões de Q, portanto, de acordo com a Tese Church-Turing, todas as funções efetivamente computáveis são representáveis em Q. Estamos particu-

[13] Para facilitar a leitura de fórmulas, quando o contexto não produzir ambiguidade entre número e numeral, vamos evitar a utilização de negrito.

[14] Podemos também dizer que uma relação *R* é representável em uma teoria sse sua função característica é representável nessa teoria.

larmente interessados em um grupo de funções, aquelas funções sintáticas necessárias para a construção de sentenças autorreferenciais.

A função sintática que iremos utilizar para a construção de autorreferência é a diagonal. A diagonal de A é $\exists x(x=\ulcorner A \urcorner \wedge A)$. Dada a codificação por números de gödel introduzida anteriormente, podemos utilizar a função numérica associada, $diag(x)$, que leva de números para números, de $gn(A)$ para $gn(\exists x(x=\ulcorner A \urcorner \wedge A))$, por exemplo. Não é difícil ver que $diag(gn(A))$ é efetivamente computável (para qualquer fórmula A). Grosso modo, o procedimento consiste em formar o número de gödel de A, o seu numeral e encaixá-los na codificação do número de gödel da diagonal de A. Portanto, existe uma fórmula, $Diag(x, y)$, na linguagem da Aritmética, tal que, $diag(n)=k$ sse $Q \vdash \forall y(Diag(n, y) \leftrightarrow y=k)$. Isso é suficiente para demonstrar o resultado que garante a existência de sentenças autorreferenciais em todas as extensões de Q, o Lema Diagonal (LD):

> **Lema Diagonal:** seja Σ uma teoria de primeira ordem que representa todas as funções recursivas e $B(x)$ uma fórmula qualquer da sua linguagem.[15] Assim, existe uma sentença G, tal que,
>
> $$\Sigma \vdash G \leftrightarrow B(\ulcorner G \urcorner)$$

Demonstração: considere a fórmula $\exists y(Diag(x, y) \wedge B(y))$, abreviadamente, F. Compute agora a diagonal de F. Obtemos assim a sentença G:

$$\exists x(x=\ulcorner F \urcorner) \wedge \exists y(Diag(x, y) \wedge B(y))$$

Grosso modo, a sentença G diz que existe uma diagonal de F e que ela satisfaz B. A diagonal de F, entretanto, é a sentença G, ela mesma. Assim, G diz de si mesma que ela é B. Classicamente, isto implica que i) $\Sigma \vdash G \leftrightarrow \exists y(Diag(\ulcorner F \urcorner, y) \wedge B(y))$. Além disso, a representabilidade de $diag$ garante que ii) $\Sigma \vdash \forall y(Diag(\ulcorner F \urcorner=y) \leftrightarrow y=\ulcorner G \urcorner)$. Logo, por i e ii,

[15] Por razões didáticas, estamos considerando uma versão mais simples do (LD) e nos restringindo às fórmulas nas quais apenas uma variável ocorre livre.

$\Sigma \vdash G \leftrightarrow \exists y(y = \ulcorner G \urcorner \wedge B(y))$. Por sua vez, isso implica classicamente que $\Sigma \vdash G \leftrightarrow B(\ulcorner G \urcorner)$.[16]

Assim, todas as teorias que sejam capazes de representar as funções recursivas incluem sentenças autorreferenciais. Como vimos, a função diagonal é efetivamente computável e, pela Tese Church-Turing, todas as funções efetivamente computáveis são recursivas. Todas as funções recursivas são representáveis em todas as extensões de Q. Logo, todas as extensões de Q incluem sentenças autorreferenciais. Veremos em breve que isto impõe importantes limites para teorias aritméticas, limites relacionados diretamente aos paradoxos apresentados na primeira parte. Antes, precisamos introduzir o Esquema T.

3.2. O Esquema T

Vamos tratar adiante de outra importante propriedade de teorias que precisa ser previamente definida, a propriedade de conter o seu próprio predicado-verdade. Grosso modo, uma teoria que contém seu próprio predicado-verdade é uma teoria que permite expressar fatos acerca da verdade ou falsidade das sentenças da própria linguagem na qual é formulada. De modo mais rigoroso, uma teoria Σ contém seu próprio predicado-verdade sse: i) Existe um predicado $T(x)$ na linguagem de Σ, tal que, para toda sentença A na linguagem de Σ, existe um termo $\ulcorner A \urcorner$ de Σ que representa um nome de A em Σ. ii) Σ acarreta todas as instâncias do Esquema T, qual seja:

$$T(\ulcorner A \urcorner) \leftrightarrow A$$

Em Tarski (1935) e Tarski (1944), o Esquema T é parte essencial da caracterização de um predicado-verdade. O Esquema T garante que o predicado em questão seja de fato um predicado-verdade, ou seja, é satisfeito por todas, e somente, as sentenças verdadeiras da linguagem. Neste sentido, dizemos ainda que o Esquema T garante a adequação material do predicado. Co-

[16] A demonstração resumida acima leva em consideração a pobreza de termos da linguagem da Aritmética. Para uma versão mais intuitiva do (LD), considere a seguinte versão análoga de função diagonal: diag*(A) = A(⌜A⌝). Considere a fórmula F seguinte: B(diag*(x)). Compute diag*(F), obtendo a sentença G, qual seja, B(diag*(⌜B(diag*(x))⌝)). Note que a sentença G diz que a diagonal de F é B, mas a diagonal de F é a sentença G, ela mesma.

mo apontado em Haack (1978, pp.144-145), dada uma linguagem L e duas definições alternativas de verdade para as sentenças de L, D_1 e D_2, se ambas forem materialmente adequadas, ambas terão exatamente a mesma extensão.[17]

Tarski obteve importantes resultados positivos na construção de definições de verdade. Tais definições estão limitadas por um resultado negativo igualmente importante, o Teorema da Indefinibilidade. Grosso modo, o teorema de Tarski impõe certos limites ao poder expressivo de linguagens, limites especialmente relacionados aos requisitos para o fechamento semântico de uma linguagem. Uma linguagem é semanticamente fechada sse ela é capaz de expressar os fatos relativos à sua própria semântica (interpretação). Uma linguagem semanticamente fechada deve conter seu próprio predicado verdade e acarretar todas as instâncias do Esquema T. Como veremos a seguir, Esquema T e (LD) são conjuntamente inconsistentes.

3.3. O Teorema da Indefinibilidade

O Teorema da Indefinibilidade (TI) foi demonstrado primeiramente por Tarski (1935). Embora Tarski não utilize a terminologia e os métodos aqui esboçados, demonstra essencialmente o mesmo resultado. A ideia é que se uma teoria é forte o suficiente para satisfazer os requisitos do (LD), então ela não pode consistentemente conter o seu próprio predicado-verdade.

> **Teorema da Indefinibilidade:** Seja Σ uma teoria de primeira ordem clássica que representa todas as funções recursivas. Assim, se Σ é consistente, Σ não pode conter o seu próprio predicado-verdade.

Demonstração: suponha que Σ é uma teoria que contém o seu próprio predicado-verdade. Assim, existe um predicado $T(x)$ na linguagem de Σ, tal que, para toda sentença A na linguagem de Σ, $\Sigma \vdash T(\ulcorner A \urcorner) \leftrightarrow A$, ou seja, Σ acarreta todas as instâncias do Esquema T. Pelo (LD), se Σ representa todas as funções recursivas e $T(x)$ é uma fórmula da sua linguagem, então, existe uma sentença G da sua linguagem, tal que, $\Sigma \vdash G \leftrightarrow \neg T(\ulcorner G \urcorner)$ (se $T(x)$ é fórmu-

[17] Tarski toma por linguagens sistemas formais, ou seja, linguagens formais interpretadas e acompanhadas de um conjunto de axiomas e regras de derivação.

la, $\neg T(x)$ é fórmula). Pelo Esquema T, $\Sigma \vdash T(\ulcorner G \urcorner) \leftrightarrow G$. Pela transitividade da bicondicional, $\Sigma \vdash T(\ulcorner G \urcorner) \leftrightarrow \neg T(\ulcorner G \urcorner)$. Logo, Σ é inconsistente.[18]

Outro modo de enunciar o Teorema de Tarski recorre à noção de definibilidade aritmética. Pelo que vimos anteriormente, o conjunto das sentenças verdadeiras da linguagem da Aritmética no modelo padrão é aritmeticamente definível sse existe uma fórmula $T(x)$ na linguagem da Aritmética, tal que, A é verdadeira no modelo padrão (ou seja, $V \vdash A$) sse $T(\ulcorner A \urcorner)$ é verdadeira no modelo padrão (ou seja, $V \vdash T(\ulcorner A \urcorner)$, pois vimos que definibilidade aritmética equivale a representabilidade em V), para qualquer sentença A da linguagem da Aritmética. Assim, o conjunto das sentenças verdadeiras no modelo padrão é aritmeticamente definível sse $V \vdash A \leftrightarrow T(\ulcorner A \urcorner)$, para toda sentença A da linguagem. Entretanto, V é uma extensão de Q, portanto representa todas as funções recursivas. Por (LD), existe uma sentença G, tal que, $V \vdash G \leftrightarrow \neg T(\ulcorner G \urcorner)$. Por redução, concluímos que o conjunto das sentenças verdadeiras da linguagem da Aritmética no modelo padrão não é aritmeticamente definível (não é representável em V). Isto significa, entre outras coisas, que V não é recursivamente axiomatizável e que, portanto, as teorias aritméticas humanamente manuseáveis são incompletas (existem sentenças aritméticas que elas não podem demonstrar nem refutar).

Como dissemos antes, (TI) espelha os mentirosos em um cenário estritamente formal. A ideia é que os mentirosos resultam de certas aplicações do (LD) em teorias semanticamente fechadas (ou seja, teorias que contenham seu próprio predicado de verdade). A sentença G introduzida na prova do teorema funciona exatamente como uma versão do Mentiroso Reforçado. Incentivamos o leitor a verificar como as outras versões seriam formalizadas.

4. Abordagens filosóficas

O Teorema da Indefinibilidade de Tarski, como vimos, não é um paradoxo, mas um resultado formal a respeito do conceito de verdade. Todavia, este teorema identifica certos limites que espelham intuitivamente os argu-

[18] Se uma teoria demonstra $A \leftrightarrow \neg A$, então ela é inconsistente. Aliás, este fato pode ser demonstrado sem pressupor o (PTE) ou qualquer outra regra não intuicionista, apenas *Consequentia Mirabilis*.

mentos anteriormente classificados como paradoxos. Não é claro, entretanto, qual seja exatamente a lição que podemos extrair do teorema de Tarski acerca dos paradoxos.

Grosso modo, o teorema de Tarski nos diz que uma teoria semanticamente fechada e clássica será também inconsistente. Nestas circunstâncias, como vimos, o Lema Diagonal estabelece a existência de sentenças mentirosas, como as sentenças (1)-(5) da segunda seção. Na presença de tais sentenças, do Esquema T e da Lógica Clássica, somos levados à inconsistência e à trivialidade.

Mesmo as teorias aritméticas mais fracas, como vimos, satisfazem os antecedentes do Lema Diagonal. Neste caso, as sentenças paradoxais funcionam como provas por redução ao absurdo da indefinibilidade, tais teorias não são semanticamente fechadas, não possuem seu próprio predicado-verdade. E quanto à linguagem natural? Podemos extrair alguma lição desse resultado formal para a linguagem natural?

A linguagem natural está contaminada de ambiguidades, vagueza e todo tipo de imprecisão que inviabiliza a sua formalização. A conclusão de Tarski neste ponto é bastante prudente:

> Nossa linguagem cotidiana não é, com certeza, uma linguagem com uma estrutura especificada com exatidão. Não sabemos precisamente quais expressões são sentenças, e sabemos em grau ainda menor quais sentenças devem ser tomadas como afirmáveis. Assim, o problema da consistência não tem um significado exato com respeito a essa linguagem. No melhor dos casos, podemos apenas arriscar o palpite de que uma linguagem cuja estrutura tenha sido exatamente especificada, e que se assemelhe à nossa linguagem cotidiana o mais possível, seria inconsistente (TARSKI, [1944], 2006, p. 170).

A linguagem natural não é uma linguagem formal, mas podemos certamente tentar construir linguagens e teorias que espelhem (ainda que parcialmente) o conceito ordinário de verdade. Como vimos, não existe uma teoria clássica e consistente que contenha o seu próprio predicado-verdade e represente todas as funções recursivas. A questão que podemos nos colocar agora é: estes requisitos são todos essenciais para o conceito ordinário de verdade? Este é o problema filosófico levantado pelos mentirosos. Diferentes tentativas de resolver o paradoxo pretendem argumentar contra um ou mais destes re-

quisitos, revisando de algum modo nossa compreensão acerca da linguagem natural e do conceito ordinário de verdade. Existe uma literatura imensa de tentativas de solução ao paradoxo do Mentiroso. Não pretendemos sequer esboçar aqui uma classificação desta literatura. No que segue iremos apenas tecer breves comentários sobre as principais famílias de soluções.[19]

As **abordagens hierárquicas** são baseadas em hierarquias geradas pelas restrições impostas pelo (TI). O predicado-verdade de uma linguagem *L* pode ser introduzido na sua metalinguagem *ML*, desde que *ML* seja essencialmente mais rica do que *L*, ou seja, desde que *L* não seja capaz de expressar *ML* em si mesma. Teremos assim, uma hierarquia de linguagens, onde o predicado verdade de cada linguagem é sempre introduzido na sua metalinguagem. Para dar um exemplo, considere a linguagem da Aritmética de primeira ordem na interpretação padrão. Assim, teremos a seguinte hierarquia:

$L_0 = \{0, s, +, \times, <\}$
$L_1 = \{0, s, +, \times, <, T_1\}$
$L_2 = \{0, s, +, \times, <, T_1, T_2\}$
$\vdots$

Nesta hierarquia, a partir do estágio 2, um predicado-verdade é introduzido em cada estágio e este funciona como o predicado verdade da linguagem cujo estágio lhe antecede. Assim, T_1 é o predicado-verdade de L_1, T_2 é o predicado-verdade de L_2, etc. A interpretação fornecida para L_1 é a interpretação padrão e, em qualquer estágio m, o predicado introduzido nesse estágio (qual seja, T_{m-1}) deve satisfazer uma versão restrita do Esquema T, ou seja, $T_{m-1}(\ulcorner A \urcorner) \leftrightarrow A$ (para qualquer sentença A de L_{m-1}).

Note que os requisitos do (LD) continuam disponíveis para todas as linguagens em todos os estágios da hierarquia. Assim, se Σ_m é o conjunto das sentenças verdadeiras da linguagem L_m, então, Σ_m é a extensão do predicado T_m em L_{m+1}. Pelo (LD), existe uma sentença G de L_m, tal que,

[19] Para uma apresentação mais detalhada das teorias aqui esboçadas, recomendamos Barrio (2014) e também Cardoso (2018). Uma apresentação mais completa (embora menos detalhada) desta literatura pode ser encontrada no verbete *Liar Paradox* da *Stanford Encyclopedia of Philosophy*, escrita por JC Beall, Michael Glanzberg e David Ripley. O texto está disponível no seguinte endereço eletrônico: https://plato.stanford.edu/entries/liar-paradox/.

$\Sigma_m \vdash \neg T_{m+1}(\ulcorner G \urcorner) \leftrightarrow G$, já que $T_{m+1}(x)$ é uma fórmula de L_m e Σ é uma extensão de Q. Pela versão restrita do Esquema T, temos que $\Sigma_m \vdash T_m(\ulcorner A \urcorner) \leftrightarrow A$, para toda sentença A de L_m. Neste caso, entretanto, a transitividade da bicondicional não parece mais implicar inconsistência, como antes.

A principal objeção que tem sido apresentada contra as soluções hierárquicas é que elas parecem impedir qualquer aproximação ou espelhamento dos aspectos essenciais da linguagem natural por parte das linguagens formais. Citando Tarski:

> Um traço característico da linguagem coloquial (ao contrário de várias linguagens científicas) é sua universalidade. Não estaria em harmonia com o espírito desta linguagem se, em alguma outra linguagem, ocorresse uma palavra que não pudesse ser nela traduzida. Poder-se-ia alegar que 'se podemos falar de modo significativo sobre alguma coisa que seja, também podemos falar a respeito dela na linguagem coloquial (TARSKI, [1935], 2006, p.32).

O poder expressivo da linguagem natural parece ser, de fato, uma de suas características essenciais. Como disse uma vez Millôr Fernandes: "se uma imagem vale mais do que mil palavras, então diga isto com uma imagem". Por outro lado, como vimos, linguagens de primeira ordem classicamente interpretadas e semanticamente fechadas são inconsistentes. Isto parece indicar que todos os modos consistentes de se preservar o poder expressivo da linguagem natural envolvem necessariamente a adoção de interpretações não clássicas.

As **abordagens não clássicas** dos paradoxos consistem em tentativas de evitar os paradoxos por meio de ajustes em relação à lógica clássica. Vamos falar muito brevemente de apenas dois tipos de soluções não clássicas, aquelas que admitem lacunas de valores-verdade (*truth-value gaps*) e aquelas que admitem aglutinações de valores-verdade (*truth value gluts*).

Os defensores de **lacunas** (*gaps*) alegam que uma teoria que espelha o conceito ordinário de verdade deve admitir que algumas sentenças não possuem valor de verdade, não são verdadeiras nem falsas. Tomemos como exemplo o Mentiroso Simples, a sentença (1) que diz de si mesma que ela é falsa. Assim, (1) é verdadeira sse (1) é falsa. O problema parece presumir que (1)

possua valor de verdade. Se recusarmos este pressuposto, o paradoxo inicialmente parece desaparecer. Como vimos, esta ideia precisa ser cuidadosamente detalhada, pois a mera adoção de lacunas resolve o Mentiroso Simples sem resolver o Mentiroso Reforçado.

Em seu trabalho especialmente importante sobre lacunas, Kripke (1975) mostrou que uma teoria de primeira ordem pode satisfazer os requisitos do (LD) e conter seu próprio predicado de verdade, desde que este predicado seja apenas parcialmente interpretado. Assim, na teoria de Kripke, T é interpretado parcialmente, a partir dos conjuntos Γ_1 (a extensão de T) e Γ_2 (a antiextensão de T). Esta interpretação deve ser consistente, ou seja, $\Gamma_1 \cap \Gamma_2$ é vazio. Além disso, se esta interpretação vai de fato garantir que a teoria em questão contenha seu próprio predicado verdade, a extensão de T deve coincidir com o conjunto de sentenças verdadeiras da linguagem da teoria e a anti extensão de T deve coincidir com o conjunto das sentenças falsas da linguagem da teoria.

Para cada par (Γ_1, Γ_2) que adotamos como interpretação de *T*, obtemos um modelo da linguagem em questão e, consequentemente, um par (Δ_1, Δ_2), correspondente aos conjuntos das sentenças verdadeiras e falsas da linguagem em questão na interpretação dada. Isto também depende do esquema de interpretação que estamos utilizando para os símbolos lógicos. Cada esquema de interpretação vai determinar um operador J definido sobre pares (Γ_1, Γ_2) de conjuntos de sentenças da linguagem. $J\ (\Gamma_1, \Gamma_2) = (\Delta_1, \Delta_2)$ sse Δ_1 é o conjunto de sentenças que resultam verdadeiras na interpretação (Γ_1, Γ_2) e Δ_2 é o conjunto das sentenças que resultam falsas na interpretação (Γ_1, Γ_2).[20] Nestes termos, o fechamento semântico, para Kripke, consiste em uma interpretação que atribua um par (Γ_1, Γ_2) a *T*, tal que $J(\Gamma_1, \Gamma_2) = (\Delta_1, \Delta_2)$. Tais interpretações são também chamadas de *pontos fixos*. (Vamos nos restringir aos pontos fixos consistentes).

[20] Mais precisamente, a interpretação ($\mathbf{\Gamma}_1$,$\mathbf{\Gamma}_2$) é a extensão do modelo padrão da linguagem em questão pela inclusão de $\mathbf{\Gamma}_1$ como extensão de *T* e $\mathbf{\Gamma}_2$ como antiextensão de *T*, de acordo com o esquema de interpretação associado a J.

Como veremos, pontos fixos consistentes terão de ser interpretações parciais. O par (Γ_1, Γ_2) é uma interpretação parcial sse $\Gamma_1 \cup \Gamma_2 \neq S$, onde *S*, é o conjunto das sentenças da linguagem em questão. Na medida em que vamos admitir interpretações parciais, precisamos adotar esquemas não clássicos de interpretação dos símbolos lógicos. Diferentes esquemas poderiam ser utilizados aqui. A única restrição é que tais esquemas devem preservar a monotonicidade sobre as interpretações de *T*. Grosso modo, isto quer dizer o seguinte: seja $J(\Gamma_1, \Gamma_2)=(\Delta_1, \Delta_2)$ e $J(\Gamma_3, \Gamma_4)=(\Delta_3, \Delta_4)$. Assim, J é um operador monotônico sse $\Gamma_1 \subseteq \Gamma_3$ e $\Gamma_2 \subseteq \Gamma_4$ implica que $\Delta_1 \subseteq \Delta_3$ e $\Delta_2 \subseteq \Delta_4$.[21]

Kripke mostrou que os operadores monotônicos possuem pontos fixos e, portanto, que existem interpretações parciais que estendem a interpretação padrão da linguagem da Aritmética para uma linguagem que, de certo modo, contém seu próprio predicado-verdade. A rigor, a existência de lacunas inviabiliza a validade irrestrita do Esquema T. Podemos, entretanto, manter o esquema na forma inferencial.[22]

Os modelos de Kripke fornecem ainda uma ferramenta técnica que permite representar importantes distinções entre as sentenças autorreferenciais exploradas na segunda seção, como as sentenças infundadas, as sentenças paradoxais, os mentirosos contingentes e as sentenças *up for grabs*. Grosso modo, as sentenças infundadas são aquelas que não pertencem ao ponto fixo mínimo. As sentenças paradoxais são aquelas que pertencem a nenhum ponto fixo (como as sentenças (1) a (5) e (9), colocadas na segunda seção). Algumas sentenças infundadas não são paradoxais, como o *truth teller* (8). Note ainda que as sentenças (6) e (7) são autorreferenciais, mas nada impede que elas se-

[21] Kripke não privilegia um esquema monotônico particular, mas apresenta algumas diferenças entre adotar três esquemas muito importantes: Kleene Forte, Kleene Fraco e as Sobrevalorações. Explorando os diferentes pontos fixos obtidos por estes diferentes esquemas, Kripke pôde traçar importantes características das sentenças mentirosas. O esquema forte de Kleene (K_3), formulado em Kleene (1952), desempenha papel especial na teoria. Para uma breve compreensão deste esquema (seu fragmento proposicional), considere as seguintes regras: i) Negação: $\neg A$ é verdadeira sse A é falsa. $\neg A$ é falsa sse A é verdadeira. De outro modo, A não tem valor. ii) $A \vee B$ é verdadeira sse A é verdadeira ou B é verdadeira. $A \vee B$ é falsa sse A e B são falsas. De outro modo, $A \vee B$ não tem valor.

[22] Afinal, se A não tem valor, então $T(\ulcorner A \urcorner)$ não tem valor. Neste caso, em k_3, $T(\ulcorner A \urcorner) \leftrightarrow A$ também não tem valor. Não obstante, M é modelo de $T(\ulcorner A \urcorner)$ sse M é modelo de A. Podemos também expressar a forma inferencial do Esquema T com as Regras T. i) Introdução de T: A implica logicamente $T(\ulcorner A \urcorner)$. ii) Eliminação de T: $T(\ulcorner A \urcorner)$ implica logicamente A.

jam incluídas na extensão de T em certos pontos fixos e isso não parece arbitrário (como o valor de (8) nos modelos em que tem valor). Finalmente, o valor de (9) depende de fatos contingentes. Se, de acordo com o modelo adotado, existe vida em Trappist-1, então (9) é infundada neste modelo. Se não existe vida em Trappist-1, (9) é falsa neste modelo.

A principal objeção frequentemente levantada às lacunas é que tais abordagens parecem deixar livre o caminho para o *Problema da Vingança*. Este problema, de certo modo, também parece denunciar perda de poder expressivo das teorias. Vejamos informalmente o argumento que leva a ele.

Suponha que Σ é uma teoria de primeira ordem que satisfaz os requisitos do (LD), é parcialmente interpretada (pelas regras de K_3) e contém um predicado T que valida as Regras T. Neste caso, (LD) garante a existência de sentenças mentirosas, mas isto não leva a paradoxos, pois estas sentenças não estão nem na extensão nem na anti extensão de *T*. Podemos ver, do lado de fora de Σ, que ela é livre de paradoxos, mas não podemos expressar isso em Σ. Se (1) não possui valor de verdade, $T(1)$ e $\neg T(1)$ também não possuem valor de verdade, então não podemos tornar verdadeira a sentença que diz que (1) não é verdadeira. Suponha que adicionássemos à linguagem de Σ o predicado $Gap(x)$ verdadeiro de todas e somente as lacunas (i.e., as sentenças nem verdadeiras nem falsas). Ora, neste caso, (LD) nos permitiria obter uma sentença G'que diz de si mesma que ela é falsa ou lacuna (aplicamos o lema à fórmula $\neg T(x) \vee Gap(x)$). Estamos de volta com os paradoxos, pois se G' é uma lacuna, então G' é verdadeira, se G' é falsa, ela é verdadeira, mas se G' é verdadeira, então, é falsa ou lacuna. Kripke (1975) reconheceu o problema da vingança como um limite para o seu próprio método, nas suas palavras:

> (...) as sentenças do Mentiroso não são verdadeiras na linguagem objeto, no sentido de que o processo indutivo nunca as torna verdadeiras; mas estamos impedidos de dizer isso na linguagem pela nossa interpretação da negação e do predicado verdade. Se nós pensarmos no ponto fixo mínimo, digamos sob a valoração de Kleene, como dando um modelo da linguagem natural, então o sentido em que nós podemos dizer, em linguagem natural, que a sentença do mentiroso não é verdadeira deve ser pensada como associada com algum estágio posterior no desenvolvimento da linguagem natural, em que o falante reflete sobre o processo de geração conduzindo ao ponto fixo mínimo. Isso não é em si mesmo parte do processo. A necessi-

> dade de ascender a uma metalinguagem pode ser uma das fraquezas da presente teoria. O fantasma da hierarquia de Tarski ainda está conosco (KRIPKE, 1975, p. 714).

Tendo isso em vista, o poder expressivo próprio da linguagem natural coloca limites em uma abordagem de lacunas para lidar com o Mentiroso. Se pretendemos, de algum modo, lidar com as características intuitivas da linguagem natural, uma abordagem de lacunas parece não ser adequada a esse programa.

Teóricos dialeteístas, como Graham Priest, sustentam que o dialeteísmo supera o problema da vingança, não sofrendo limitações de expressividade. O custo para conquistar o poder expressivo estaria justamente em assumir **aglutinações** (*gluts*).

Dialeteísmo é a visão de que há contradições verdadeiras (dialeteias). Na perspectiva dialeteísta, a resiliência do Mentiroso diante dos tratamentos consistentes estaria mostrando justamente que há contradições verdadeiras e que, portanto, o dialeteísmo é o tratamento adequado para o paradoxo. Nessa abordagem, a sentença do mentiroso resulta verdadeira e falsa (uma aglutinação). Para defender esse ponto, normalmente é empregado o caráter semanticamente fechado da linguagem natural e a alegação de que a maioria das tentativas de impedir a contradição viola algum recurso intuitivo das línguas naturais que leva ao paradoxo. Como o paradoxo surge na linguagem natural, a partir de recursos básicos, tratar o Mentiroso, na perspectiva dialeteísta, é assumir que há contradições verdadeiras.

> De maneira geral, tais paradoxos como o Mentiroso fornecem alguma evidência para a tese dialeteísta que algumas contradições são demonstravelmente verdadeiras, no sentido que elas são implicadas por fatos manifestos sobre as línguas naturais e nossos processos de pensamentos. Os paradoxos do Mentiroso estendido como 'Esta sentença não é verdadeira' são expressáveis no português ordinario. Suas características paradoxais, como os dialeteístas salientam, são devidas justamente às características intuitivas da linguagem ordinária (PRIEST, BERTO, WEBER, 2018, p. 20).

Segundo Priest (2006, p. 24), a consistência da linguagem natural só pode ser adquirida ao diminuir a sua capacidade expressiva de algum modo.

Assim, com base no fenômeno da vingança, um argumento a favor do dialeteísmo pode ser estabelecido através do seguinte dilema:

> **D1**: qualquer teoria da verdade em uma determinada língua natural será incompleta ou aglutinada (*glutty*).

O dilema colocado por meio da disjunção é demonstrado a partir do próprio Mentiroso. Ou evitamos a força expressiva da linguagem, impedindo o Mentiroso (e, portanto, a aglutinação), ou permitimos toda expressividade e derivamos o Mentiroso. No primeiro caso, a teoria é incompleta em sua força expressiva e a disjunção vale. No segundo caso, ao permitir toda capacidade expressiva da linguagem, temos meios para construir o Mentiroso e, assim, a disjunção também vale. Dessa forma, de acordo com os argumentos de Priest, devemos aceitar o dialeteísmo. Na visão de Priest, a completude expressiva é adquirida por um preço não tão difícil de pagar: aglutinações. Levando em consideração esse argumento, a teoria de aglutinações decorre da promessa de completude expressiva.

O dialeteísmo afirma que algumas contradições são verdadeiras, mas não que tudo é verdadeiro. Assim, uma lógica paraconsistente deve ser adotada para evitar a trivialidade, restringindo o (PE). Priest (1979) concebe uma lógica, intitulada *Lógica do Paradoxo* (LP) que, segundo ele, é adequada para modelar contradições verdadeiras, como a do Mentiroso, sem trivialidade. LP tem sentenças que são "somente verdadeiras", "somente falsas" e "verdadeiras e falsas" (aglutinações). A atribuição de valores de verdade é dada a partir de uma função do conjunto das fórmulas atômicas no conjunto de valores de verdade, e é estendida para todas as fórmulas a partir do comportamento dos conectivos dados pelas tabelas de K_3. Diferentemente de K_3, onde o terceiro valor é interpretado como uma lacuna, em LP o terceiro valor é interpretado como uma aglutinação. Outra diferença é que em LP o conjunto designado (valores que são preservados em inferências válidas) contém o valor aglutinado, além do verdadeiro.

Em LP, o princípio de explosão (PE) ($A \wedge \neg A \vDash B$) não vale irrestritamente. Há modelos que satisfazem $A \wedge \neg A$, mas não satisfazem *B*. Isso

ocorre quando o valor de A é aglutinado (que é designado) e o valor de B é falso. Note que aqui enunciamos (PE) na forma inferencial e não na forma condicional. Em virtude do TD, clássicamente, todas as verdades lógicas na forma condicional são também teoremas e vice versa. Em LP, a condicional também pode ser definida por meio da equivalência clássica entre $A \rightarrow B$ e $\neg A \vee B$. Assim, o mesmo contra modelo anteriormente apresentado para a forma inferencial de (PE) também invalida (MP) e a forma condicional de (PE) (cf. PRIEST, 2008, cap. 7).

A partir daí, Priest considera que, por meio de aglutinações é possível modelar a contradição do Mentiroso em LP e ainda preservar a força expressiva da linguagem da teoria. Aparentemente, para os que estão dispostos a assumir que há contradições verdadeiras, o dialeteísmo traria algumas vantagens, sobretudo se considerarmos o poder expressivo da linguagem. Todavia, mesmo assumindo contradições verdadeiras, os paradoxos colocam limites para a abordagem dialeteísta. Priest argumenta que o Mentiroso captura a noção de contraditoriedade do quadrado de oposições de Aristóteles. Entretato, se considerarmos o resultado do Mentiroso em LP temos que tanto a sentença do mentiroso quanto a sua negação resultam verdadeiras. Considerando isso, a negação de LP retrataria uma relação de subcontrariedade, como foi apontado por Slater (1995). Ora, se o Mentiroso é entendido como uma contradição (com o sentido de contraditoriedade) e LP tem uma negação com um sentido mais fraco (a subcontrariedade), então LP não seria a lógica adequada para lidar com a contradição do Mentiroso, sem trivialidade. Isso romperia com a promessa dialeteísta de lidar com o Mentiroso. Arenhart e Melo (2017) argumentam que o sentido de contradição como contraditoriedade, encapsulado no Mentiroso, leva à trivialidade, assumindo as próprias exigências do dialeteísmo. Assumindo essa linha de argumentação, a linguagem da teoria dialeteísta não seria capaz de expressar o Mentiroso sem trivialidade.

O paradoxo de Curry também apresenta desafios para a abordagem dialeteísta. Mesmo concedendo que LP captura o sentido de contradição adequado (o que é discutível), lembramos que a estratégia em LP para evitar a trivialidade funciona com base na contradição: tome uma interpretação em

que $A \wedge \neg A$ é designada e B não designada. Todavia, essa estratégia não impede a trivialidade via paradoxo de Curry que, como vimos anteriormente, leva à trivialidade diretamente, sem a necessidade da contradição. Assim, mesmo invalidando a explosão, o paradoxo de Curry não seria evitado e poderíamos derivar a trivialidade. Priest (2006) lida com esse problema restringindo a regra de contração do condicional (Con).[23]

Todavia, um outro paradoxo do tipo Curry foi colocado, desafiando os limites do dialeteísmo, a partir da noção de validade (conhecido como *Curry da validade*, ou v-Curry). Uma restrição na condicional material, como foi feita por Priest (2006), não seria suficiente para impedir a trivialidade a partir de v-Curry. Dada uma linguagem semanticamente completa e uma teoria Σ nesta linguagem dotada de uma relação de consequência $\vdash$, podemos introduzir um predicado de validade para essa linguagem $val(\ulcorner A \urcorner, \ulcorner B \urcorner)$ para expressar que B se segue de A em Σ. Com isso, é possível construir uma sentença C (conhecida como sentença Beall-Murzi) que significa $val(\ulcorner C \urcorner, \ulcorner \bot \urcorner)$, onde $\bot$ significa "tudo é verdadeiro". A partir de princípios muito razoáveis sobre a validade, Beall (2015) mostra como podemos chegar na trivialidade a partir da sentença C.

Assumindo que a linguagem é completa em sua força expressiva, ela irá fornecer ingredientes para a derivação da trivialidade e, se estamos seguindo a mesma linha de raciocínio que Priest estabeleceu para o caso do Mentiroso, (i.e., o argumento da completude expressiva via Mentiroso), devemos endossar a trivialidade, se quisermos manter a completude expressiva e os recursos que levam ao paradoxo disponíveis na linguagem natural. Diante disso, Beall (2015, p.558) estabelece um segundo dilema da seguinte maneira:

> **D2**: qualquer teoria da verdade em uma determinada língua natural será incompleta ou trivial.

Agora, se quisermos evitar a trivialidade, ou devemos evitar a completude expressiva ou algum princípio da linguagem natural que conduz ao para-

[23] Como dissemos, LP invalida (MP), portanto o paradoxo de Curry não é um problema para LP. Entretanto, uma condicional que não valide (MP) está muito longe de capturar as intuições da condicional da linguagem natural. Entretanto, Priest (2006) apresenta uma condicional intensional que valida (MP) e invalida (Con).

doxo. Ora, mas segundo o argumento a favor do dialeteísmo que vimos acima (com base em **D1**) devemos preservar a completude expressiva e os demais princípios que levam ao paradoxo (foi esse o argumento dado para aceitarmos aglutinações via paradoxo do Mentiroso). Beall chama **D1** e **D2** de "disjunções limitativas". Assim, diante de **D1** e **D2**, Beall (2015, p. 575) diz: "[p]aridade de raciocínio demanda igual tratamento de ambas disjunções limitativas (e a conclusão tirada daí). Mas, então, devemos escolher a incompletude".

Para evitar a trivialidade, devemos evitar a completude. Com isso, aglutinações perdem a motivação. Aglutinações foram motivadas pela promessa de completude. Agora, sem completude, não temos trivialidade (por conta do v-Curry), mas também não temos aglutinações; isto é, sem sentenças do Mentiroso. Ou seja, se perdemos a completude, perdemos o motivo para aceitar aglutinações. Com isso, o dilema entre completude expressiva e trivialidade parece assombrar o dialeteísmo.

Priest tenta lidar com a situação do v-Curry a partir de um sistema chamado "*Dual Bunch Logic*" em que ele estabelece dois tipos de conjunções: uma conjunção extensional e um operador de fusão. Não vamos entrar nos detalhes técnicos aqui (para isso, veja PRIEST, 2015). Com esse formalismo, a trivialidade é evitada, mas o operador de fusão colocado para resolver o problema tem pouca conexão com o comportamento da conjunção em línguas naturais. Priest diz que a fusão é um tipo de conjunção que representa uma aplicação funcional que não é nem associativa, nem comutativa e nem contrativa. Porém, Priest não fornece nenhuma relação do operador de fusão com os princípios presentes nas línguas naturais – algo que esperado pelo um dialeteísta, tendo em vista os argumentos para o dialeteísmo via recursos das línguas naturais (cf. PRIEST, 2006, p.09). Além disso, Priest evita a trivialidade com base no seu formalismo, porém ele não oferece nenhuma motivação filosófica ou razão independente para restringir a inferência do v-Curry, além do motivo de que v-Curry leva à trivialidade. Sobre oferecer uma razão independente para restringir o tipo de contração envolvida no v-Curry, Priest (2016, p.94) apenas remete aos trabalhos dos lógicos subestruturais, dizendo que nesses trabalhos podemos encontrar, dentre outras coisas, justificações filosóficas

para se restringir a contração e, com isso, impedir a derivação do v-Curry. Ocorre que os motivos que poderiam levar à falha do v-Curry, podem levar também à falha do Mentiroso, como bem apontam "os lógicos subestruturais" referidos por Priest (veja, por exemplo, ZARDINI, 2011).

Essa discussão sobre v-Curry no contexto das lógicas subestruturais é vasta e com várias implicações.[24] A partir dessa discussão entre Beall e Priest sobre a validade do v-Curry, Arenhart e Melo (2019) sustentam que os critérios de Priest são alterados a depender das consequências que ocorrem no dialeteísmo. Quando é o Mentiroso que está em jogo, então os recursos que levam ao paradoxos devem estar todos autorizados, sob pena de não termos uma solução natural para o problema. Assim, o paradoxo está autorizado e devemos aceitar o seu resultado: aglutinações. Agora, quando temos um paradoxo que leva à trivialidade diretamente, o critério é modificado, podendo assumir um formalismo que restringe os princípios das línguas naturais que levam ao seu resultado: a trivialidade.

É interessante mencionar que embora Beall se posicione contrariamente ao argumento da completude semântica em favor do dialeteísmo, ele ainda aceita o Mentiroso como uma motivação para aglutinações. Porém, a motivação de Beall para tomar o Mentiroso como uma evidência em favor de aglutinações não consiste em um argumento lógico, e sim em uma consideração estética. A conclusão de Beall é a seguinte:

> Devemos rejeitar o argumento alvo a partir da versão da inconsistência-absoluta da disjunção não limitativa e o objetivo da completude para aglutinações. Mesmo se, contrariando as aparências, a paridade de razões não exigir uma resposta igual para um perfeito paralelo da disjunção limitativa, a versões da inconsistência-absoluta da disjunção limitativa põe fim em qualquer promessa de uma teoria completa do português no português; e então a promessa de ter uma tal teoria ao mero custo de aglutinações é uma falsa promessa (BEALL, 2015, p. 580).

Depois dessa conclusão, Beall (2015, p.583) apenas diz que "o tratamento de aglutinações surge a partir de considerações estéticas: tais Mentirosos simplesmente parecem aglutinações". Porém, consideramos que esse

[24] Para uma apresentação ampla dos tratamentos subestruturais, ver BEALL, GLANZBERG,& RIPLEY, 2018, cap. 7).

argumento estético é insuficiente para sustentar o dialeteísmo (para detalhes, ver MELO, 2017). Independentemente da viabilidade do argumento estético, o dialeteísmo padece do mesmo problema de seus rivais consistentes: a falta de força expressiva da teoria e isso é reconhecido mesmo por alguns dialeteístas, como Beall.

Considerações finais

Como vimos ao longo do artigo, o paradoxo do Mentiroso é um paradoxo importante para a lógica, tendo em vista os resultados técnicos relacionados com tal paradoxo, e para a filosofia da lógica, considerando as abordagens filosóficas sobre o conceito de verdade que surgiram a partir de reflexões sobre o paradoxo do Mentiroso. Neste artigo procuramos fazer uma apresentação sobre o paradoxo do Mentiroso, destacando alguns desses resultados técnicos e filosóficos relacionados com ele. Além disso, procuramos deixar claro que o paradoxo do Mentiroso parece impor limites que são difíceis de acomodar em teorias que pretendem dar conta do predicado-verdade ordinário.

Começamos este texto apresentando o paradoxo do Mentiroso para deixar claro o problema mesmo para os não familiarizados com as discussões sobre paradoxos e teorias da verdade. A partir dessa apresentação foi possível conferir também alguns pressupostos importantes envolvidos no Mentiroso. Como tentamos mostrar, existem versões naturais dos mentirosos apoiadas em premissas que espelham traços aparentemente ineliminâveis da linguagem natural e do conceito ordinário de verdade. Tais versões nos colocam diante de um problema filosófico genuíno. Em primeiro lugar, a existência de sentenças autorreferenciais não parece ser considerada consensualmente condição necessária nem suficiente para a derivação dos paradoxos. Além disso, mesmo em algumas teorias formais (aquelas capazes de representar todas as funções recursivas) existem sentenças que espelham a autorreferência. Isto nos levou ao Teorema da Indefinibilidade de Tarski, pelo qual, nenhuma teoria clássica capaz de representar todas as funções recursivas pode (consistentemente) conter o seu próprio predicado-verdade.

Em segundo lugar, o conceito ordinário de verdade (aquele representado pelo predicado-verdade da linguagem natural) parece satisfazer o Esquema T e é difícil encontrar bons argumentos contra o fechamento semântico da linguagem natural. Ou seja, diferentemente do que dissemos acima sobre certas linguagens e teorias formais, parece difícil recusar que a linguagem natural contenha o seu próprio predicado-verdade. Finalmente, presumimos que o conceito ordinário de verdade seja consistente, não trivial e compatível com a lógica clássica. Ou seja, presumimos que uma teoria da verdade que represente o conceito ordinário de verdade deveria ser uma teoria clássica, consistente e não trivial. Como vimos, as principais tentativas de solução, ainda que avancem importantes resultados para uma melhor compreensão dos mentirosos e dos limites estabelecidos pelo Teorema de Tarski, enfrentam problemas muito semelhantes.

Nas abordagens hierárquicas e nas abordagens com lacunas temos de escolher entre perder poder expressivo ou abrir o caminho para versões mais fortes do Mentiroso. Neste caso, tentamos ao menos preservar a consistência, mas temos de pagar com a perda de expressividade.

De outro lado, as abordagens com aglutinação invertem essa relação, abrem mão da consistência pela recompensa de poder expressivo. Não é claro, entretanto, que a recompensa seja obtida. Como expresso pelos argumentos de Beall (2015), via a paridade de D1 e D2, qualquer teoria da verdade em uma determinada língua natural será expressivamente incompleta ou trivial. Assim, considerando os argumentos de Beall, as abordagens dialeteístas que se pretendam não triviais serão também assombrada pelo fantasma de Tarski, na medida em que não podem pagar a promessa da completude expressiva, mesmo assumindo aglutinações.

Considerando os argumentos que foram desenvolvidos por Arenhart e Melo (2019), indicamos aqui que as abordagens de aglutinações até podem se livrar do fantasma de Tarski, mas, para tanto, terão que enfrentar o fantasma da trivialidade. A contradição envolvida no Mentiroso parece impor limites que variam entre a expressividade e a trivialidade. Esse dilema nos permite afirmar que o paradoxo do Mentiroso diz respeito a um problema genuina-

mente filosófico que pode pode ser acomodado por nada menos do que um repúdio de parte da nossa herança conceitual, como observado por Quine (1966, p.11).

Referências

ARENHART, J. R. B. & MELO, E. S. 'Classical Negation Strikes Back: Why Priest's Attack on Classical Negation Can't Succeed'. *Logica Universalis* (Print), v. 11, 2017, pp. 465 - 487.

———— '(De)motivating Gluts'. *Logique et Analyse*, v. 246, 2019, p p. 115 -134.

BARRIO, E. *La Lógica de la Verdad*. Eudeba Universidad de Buenos Aires, 2014.

BARWISE, J., AND ETCHEMENDY, J. *The Liar*: an essay on truth and circularity. Oxford, New York, 1987.

BEALL, J.C *Revenge of The Liar - New Essays on the Paradox*. OUP, New York, 2007.

————'Trivializing sentences and the promise of semantic completeness'. *Analysis* 75(4), 2015 pp.573 – 584.

BEALL, J.C., GLANZBERG, M. & RIPLEY, D. *Formal theories of truth*. Oxford: Oxford Un. Press, 2018.

BOOLOS, G., BURGESS, J., AND JEFFREY, R. *Computabilidade e Lógica* [2007]- Tradução de Cezar Mortari. Unesp, São Paulo, 2012.

BURGE, T. 'Semantical paradox'. *Journal of Philosophy*. 76, 1979, pp. 169 – 198.

CARDOSO, G. *O Paradoxo do Mentiroso: uma introdução*. Coleção CLE, 2018.

FRANKFURT, H. *Sobre falar merda*. Ed Intrínseca Limitada, 2005.

GLANZBERG, M. 'The liar in context'. *Philosophical Studies*. 103(3). 2001, pp. 217 – 251.

GLANZBERG, M. 'A contextual-hierarchical approach to truth and the liar paradox'. *Journal of Philosophical Logic*. 33(1), 2004. pp. 27 – 88.

GRIM, P. *The Incomplete Universe: Totality, Knowledge, and Truth*. MIT Press, 1991.

HAACK, S. *Filosofia das Lógicas* [1978]. Trad. Cezar Mortari e Luiz Henrique Dutra. Editora Unesp, 2002.

HECK, R. 'Self-reference and the languages of arithmetic'. *Philosophia Mathematica* 15(1), 2007, pp. 1 – 29.

KLEENE, S. *Introduction to Metamathematics*. North Holland Publishing CO., Amsterdam, 1952.

KRIPKE, S. 'Outline of a theory of truth'. *Journal of Philosophy*. 72, 1975, pp. 690 – 716.

MELO, E. S. *O paradoxo do Mentiroso: argumentos contra o dialeteísmo*. (tese de doutorado). Programa de Pós-Graduação em Filosofia, UFSC. Florianópolis, 2017.

MORTARI, C. *Introdução à Lógica*. Editora Unesp, 2016.

PARSONS, C. Parsons, C. 'The Liar Paradox". *Journal of Philosophical Logic*. 3(4), 1974, pp. 381 – 412.

PRIEST, G. 'The logic of paradox'. *Journal of philosophical logic*. 8, 1979. pp. 219 – 241

———— *In Contradiction*. Oxford University Press, Oxford, 2006.

———— *An Introduction to Non-classical Logic*. Cambridge University Press, Cambridge, 2008.

———— 'Fusion and confusion'. *Topoi* 34, 2015, pp. 55 – 61.

———— 'Old wine in (somewhat leaky) new bottles. Some comments on Beall'. *Australasian Journal of Logic*. 13(5), 2016, pp. 89 – 95

PRIEST, G., BERTO, F. and WEBER, Z., 'Dialetheism', *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/dialetheism/>.

QUINE, W. *The Ways of Paradox and Other Essays*. Random House, New York, 1966.

READ, S. [1995] *Repensando a lógica: uma introdução à filosofia da lógica* Tradução de Abílio Rodrigues. Belo Horizonte: Editora da UFMG, 2016.

RUSSELL, B. 'Mathematical logic as based on the theory of types'. *American Journal of Mathematics*. 30, 1908, pp. 222 – 262.

SAINSBURY, R. M. *Paradoxes*. Cambridge University Press, 2009.

SIMMONS, K. *Universality and The Liar: An Essay on Truth and the Diagonal Argument*. Cambridge University Press, New York, 1993.

SIMMONS, K. *Revenge and context*. In BEALL (2007), pp. 345 – 367.

SLATER, B. H. 'Paraconsistent logics?'. *Journal of Philosophical Logic*. 24. 1995, 451– 454.

SMULLYAN, R. *What is the name of this book?* Prentice-Hall Inc, New Jersey, 1978.

TARSKI, A. [1935] 'O conceito de verdade nas linguagens formalizadas' - Tradução de Cezar Mortari. In *A Concepção Semântica da Verdade - Textos Clássicos de Tarski*, [Orgs.: Cezar Mortari; Luiz Dutra]. Unesp, São Paulo, 2006, pp. 19 – 148.

———— [1944] 'A concepção semântica da verdade e os fundamentos da semântica'. Tradução de Luiz Dutra. In *A Concepção Semântica da Verdade - Textos Clássicos de Tarski*, [Orgs.: Cezar Mortari; Luiz Dutra]. Unesp, São Paulo, 2006, pp. 157 – 201.

———— [1969] 'Verdade e demonstração'. Tradução de Jesus de Paula Assis In *A Concepção Semântica da Verdade - Textos Clássicos de Tarski*, [Orgs.: Cezar Mortari; Luiz Dutra]. Unesp, São Paulo, 2006 pp. 203 – 233.

TARSKI, A., MOSTOWSKI, A., AND ROBINSON, R. *Undecidable Theories*. North Holland Publishing Company, Amsterdam, 1953.

VAN FRAASSEN, B. C. 'Pressuposition, implication and self-reference'. *Journal of Philosophy* 65(5) (1968), pp. 136–152.

YABLO, S. 'Paradox without self-reference'. *Analysis*. 53-4, 1993, pp. 251 – 252.

ZARDINI, E. 'Truth without contra(di)ction'. *The Review of Symbolic Logic*. 4(4), 2011, pp. 498 – 535.