

Revista de Pesquisa em Filosofia

FUNDAMENTO

Ouro Preto - Minas Gerais - Brasil



UFOP

**Universidade Federal
de Ouro Preto**

Fundamento Revista de Pesquisa em Filosofia / Universidade Federal de Ouro Preto / Grupo de Pesquisa em Filosofia Contemporânea. n. 12, (Jan–jun. 2016) – Ouro Preto:

Ed. UFOP, 2016

Semestral – Tiragem: 200 exemplares.

ISSN: 2177–6563

1. Filosofia - Periódicos. I. Universidade Federal de Ouro Preto. II. Instituto de Filosofia, Artes e Cultura. III. Grupo de Pesquisa em Filosofia Contemporânea.

Endereço para correspondência:

Revista Fundamento
Instituto de Filosofia, Artes e Cultura – IFAC
Rua Coronel Alves, 55, Centro, Ouro Preto, MG, Brasil
CEP: 35400-000

Endereço eletrônico: revistafundamento@ufop.br

Revista Fundamento

UNIVERSIDADE FEDERAL DE OURO PRETO

Reitor

Prof. Dr. Marcone Jamilson Freitas Souza

INSTITUTO DE FILOSOFIA, ARTES E CULTURA/IFAC

Diretor

Prof. Dr. Guilherme Paoliello

DEPARTAMENTO DE FILOSOFIA/DEFIL

Chefe de Departamento

Prof. Dr. Olímpio José Pimenta

Bibliotecária

Neide Nativa

FUNDAMENTO

REVISTA DE PESQUISA EM FILOSOFIA

Diretores Gerais

Prof. Dr. Mário Nogueira (UFOP)

Prof. Dr. Sérgio Miranda (UFOP)

Diretores Executivos

Aline Monteiro Homssi (UFOP)

Editores

Prof. Dr. Mário Nogueira (UFOP)

Prof. Dr. Sérgio Miranda (UFOP)

Prof. Me. Luiz Helvécio Marques Segundo (UFSC / UFOP)

Editor 12ª Edição

Prof. Me. Sagid Salles (UFAM)

Diagramação

Aline Monteiro Homssi (UFOP)

Assistente de Revisão

Aline Monteiro Homssi (UFOP)

Comitê Editorial

Prof. Dr. Alexandre Noronha Machado (UFPR)

Prof. Dr. Hélio Lopes da Silva (UFOP)

Prof. Dr. Danilo Marcondes (PUC-Rio)

Prof. Dr. Frank Thomas Sautter (UFSM)

Prof. Dr. Guido Imaguire (UFRJ)

Prof. Dr. Marco C. Ruffino (Unicamp)

Prof. Dr. Nelson Gonçalves Gomes (UNB)

Prof. Dr. Túlio Xavier de Aguiar (UFMG)

SUMÁRIO

9 APRESENTAÇÃO

I - ARTIGOS INÉDITOS

- 13 **REAÇÕES AO Ceticismo Radical**
SOME REACTIONS TO RADICAL SKEPTICISM
L. H. Marques Segundo - Universidade Federal de Santa Catarina / Universidade Federal de Ouro Preto
- 39 *O ENIGMA DE FREGE*
FREGE'S PUZZLE
Juliana Faccio Lima - University of California Santa Barbara
- 75 **UMA BREVE INTRODUÇÃO AO PARADOXO DA CONHECIBILIDADE**
A BRIEF INTRODUCTION TO THE KNOWABILITY PARADOX
Iago Bozza Francisco - University of Miami
- 111 **ALGUMAS CONSIDERAÇÕES SOBRE O PARADOXO DE CURRY**
SOME REMARKS ON CURRY'S PARADOX
André Nascimento Pontes - Universidade Federal do Amazonas
- 127 **O PARADOXO INFERENCIAL DE LEWIS CARROLL**
LEWIS CARROLL'S INFERENTIAL PARADOX
Rodrigo Reis Lastra Cid - Faculdade Arquidiocesana de Mariana
- 139 **O PROBLEMA DA VAGUEZA**
THE PROBLEM OF VAGUENESS
Sagid Salles - Universidade Federal do Amazonas / Programa de Pós-Graduação Lógica e Metafísica da UFRJ

Apresentação

Essa edição é dedicada aos puzzles/quebra-cabeças e paradoxos da filosofia. Um puzzle da filosofia é um problema filosófico que envolve um conjunto de crenças que, apesar de aparentemente verdadeiras, são inconsistentes entre si. Não é preciso muito para perceber porque são tão interessantes e exigem solução. Afinal, eles envolvem “crenças que, apesar de aparentemente verdadeiras, são inconsistentes entre si”. Como é evidente, estou meramente repetindo uma parte da definição informal de *puzzle filosófico* dada acima. Mas penso que os quebra-cabeças filosóficos são interessantes por si mesmos, e a mera explicação do que são já é suficiente para revelar seu interesse.

Paradoxos são argumentos aparentemente válidos, com premissas aparentemente verdadeiras e conclusão aparentemente falsa. Assim como no caso dos puzzles, é fácil perceber o seu interesse. Um argumento válido com todas as premissas verdadeiras não pode ter uma conclusão falsa. A presença de um argumento com todas essas características indica que algo está errado com alguma crença ou regra de inferência aceita. Isso é, por si só, interessante.

Por fim, paradoxos e puzzles filosóficos são problemas envolvendo conceitos muito básicos como *conhecimento, significado, referência, implicação, contradição*, etc. Puzzles e paradoxos filosóficos colocam em xeque crenças ou teorias filosóficas aparentemente verdadeiras sobre conceitos como esses, e podem nos levar – e frequentemente levam – à refutação das mesmas, assim como à formulação de novas e melhores teorias. Desse modo, eles não apenas têm interesse em si, mas também nos auxiliam a avaliar e revisar nossas crenças e teorias filosóficas.

O leitor poderá encontrar uma instância disso em cada um dos artigos publicados nesta edição. Cada um deles discute, de forma acessível, um puzzle ou paradoxo da filosofia. A edição abre com um artigo de L. H. Marques Segundo, que apresenta o puzzle cético e algumas das principais respostas ao mesmo. O segundo artigo, de Juliana Faccio Lima, é sobre o puzzle de Frege, uma discussão que certamente já se tornou clássica na filosofia

analítica. O terceiro artigo, de Iago Bozza Francisco, é uma introdução ao *paradoxo da conhecibilidade*, escrita em linguagem acessível e incluindo uma discussão mais extensa da solução de Jonathan Kvanvig. André Nascimento Pontes é o autor do quarto artigo, uma discussão muito clara do *paradoxo de Curry*, e sua relação com alguns conceitos importantes em filosofia da lógica. Rodrigo Reis Lastra Cid apresenta e discute, no quinto artigo, o paradoxo de Lewis Carroll sobre a justificação de princípios lógicos. O último artigo é de minha autoria e apresenta e discute o problema (e não as soluções) da vagueza e o paradoxo sorites. Uma boa leitura a todos.

Sagid Salles

Manaus, 2017

Artigos Inéditos

Reações ao ceticismo radical

Some reactions to radical skepticism

L. H. Marques Segundo
PPG-Fil /UFSC e Dep. de filosofia/ UFOP

*Se o cético não existisse, o
epistemólogo precisaria inventá-lo.*

Laurence Bonjour¹

Resumo

Enfrentar o desafio cético tem sido por muito tempo uma das principais motivações dos epistemólogos. Essa fixação se deve à força do argumento cético, que repousa em premissas plausíveis, mas estabelece uma conclusão inaceitável – a de que nada podemos saber sobre o mundo externo. O objetivo deste artigo é apresentar algumas das principais reações ao desafio cético que têm recebido especial atenção no debate contemporâneo.

Palavras-chave

Desafio cético. Argumento cético. Mundo externo.

1 - The Structure of Empirical Knowledge (1985).

Abstract

One of the main epistemologists' motivations since always is to rebut the skeptical challenge. This resoluteness is due to the strength of skeptical argument, which rests on plausible premises but supports an unacceptable conclusion – that we cannot know anything about the external world. The aim of this paper is to introduce some of the main contemporary responses to the skeptical challenge.

Keywords

Skeptical challenge. Skeptical argument. External world.

1. O puzzle cético

Grande parte dos problemas filosóficos surge como *puzzles* intelectuais. Um *puzzle* intelectual surge quando nos apercebemos de que sustentamos um conjunto inconsistente de crenças no qual cada uma dessas crenças, individualmente, nos parece plausível. Uma tensão é então gerada: temos de abandonar uma dessas crenças, muito embora não sejamos capazes de decidir qual. Como bem notou Bertrand Russell numa passagem memorável, “o peculiar da filosofia é começar com algo tão simples que pareça não merecer explicação, e terminar com algo tão paradoxal que ninguém acredite.”²

O *puzzle* cético é um desses típicos problemas filosóficos. Considere as seguintes afirmações:

- (1) Sei que tenho mãos.
- (2) Se sei que tenho mãos, então sei que não sou um cérebro numa cuba.
- (3) Não sei que não sou um cérebro numa cuba.

2 - *The Philosophy of Logical Atomism* (1972).

A afirmação (1) é uma dessas crenças simples que não merecem explicação; parece-nos obviamente verdadeira.

A afirmação (3), embora não tão óbvia assim, parece se mostrar verdadeira logo que pensamos sobre ela. Um pequeno conto de ficção científica nos ajuda a ver isso. Imagine que um cientista maluco remova o seu cérebro e o conserve numa cuba cheia de nutrientes. Ele então liga seu cérebro a um supercomputador e passa a mandar estímulos elétricos de modo a simular todas as suas experiências perceptuais. Tudo o que você vê, sente, ouve, etc. (incluindo este texto que você lê agora!) não é produto da interação da sua mente com os objetos externos a ela, mas sim de uma sofisticada simulação de computador.³ Essa é claramente uma possibilidade física,⁴ ainda que maluca. Como você afastaria tal possibilidade e diria que *sabe* que não é um cérebro numa cuba? Uma vez que parece não haver meios *a priori* de afastar tal possibilidade, você teria de recorrer a meios empíricos: seu amigo lhe disse que você tem um corpo (e que não tem qualquer coisa ligada ao seu cérebro), você se olhou no espelho e viu seu próprio corpo, você olhou para suas mãos, sentiu-as, e assim por diante. Ora, mas todas essas experiências poderiam ser produto de uma simulação, de modo que você não tem garantia de que sejam experiências produzidas pelo mundo externo. Inclusive, a sua crença sobre a impossibilidade tecnológica do incubamento de cérebros é produto da simulação. Se esses são os únicos modos de se saber que não somos cérebros em cubas, então não podemos saber que não somos.

Quanto à afirmação (2), o seguinte raciocínio nos mostra sua plausibilidade. Se ela fosse falsa, a sua negação teria de ser verdadeira; isto é, a proposição <Sei que tenho mãos e não sei que não sou um cérebro numa cuba> teria de ser verdadeira. Ora, a proposição <Tenho mãos>, implica logicamente a proposição <Não sou um cérebro numa cuba>. Pois no caso em que sou um cérebro numa cuba, não tenho corpo, e, portanto, não tenho mãos. Assim, se

3 - A hipótese do cérebro numa cuba, que foi pela primeira vez apresentada por Hilary Putnam (1981), é uma versão moderna da hipótese do Gênio Maligno de Descartes, apresentada em sua primeira *Meditação*. Uma apresentação sucinta e divertida pode ser encontrada em Pollock e Cruz (1999).

4 - Dizer que algo é uma possibilidade física significa dizer que nada nas leis da física impede que tal situação venha a acontecer. A possibilidade física é diferente da possibilidade lógica. Algo é logicamente possível se não implica contradição. Por exemplo, é logicamente possível que eu tenha um soco tão forte que derrube toda a Muralha da China: a afirmação *L. H. derrubou toda a Muralha da China com um soco* não implica contradição. Contudo, é uma impossibilidade física; a força de um soco humano não é suficiente (nem com toda tecnologia imaginável) para abalar e derrubar algo como a Muralha da China. A hipótese do cérebro numa cuba não é apenas logicamente, mas também fisicamente possível. Alguns leitores podem ter sido levados a pensar que tal hipótese é fisicamente impossível, uma vez que não dispomos de meios para manter um cérebro vivo dentro de uma cuba, e muito menos um supercomputador que simule a realidade externa. Mas isso só mostra que não há tecnologia para tal; essa tecnologia pode ser desenvolvida (não importa quando!), pois nada há nas leis da natureza que impeçam tal coisa. A acusação de impossibilidade, portanto, tem de se restringir à *impossibilidade tecnológica atual*, o que não implica impossibilidade física.

sei que tenho mãos, e ter mãos implica que não sou um cérebro numa cuba, como pode ser verdade que sei que tenho mãos mas não sei que não sou um cérebro numa cuba? Sustentar a negação de (2) é assentir ao que DeRose chamou de “conjunção abominável”. Isso parece ser razão suficiente para considerarmos (2) como bastante plausível.

Embora plausíveis, as afirmações (1)-(3) são inconsistentes, *i.e.* não podem ser todas verdadeiras ao mesmo tempo, sob pena de contradição. Por exemplo, (1) e (2) implicam (por *modus ponens*)

(4) Sei que não sou um cérebro numa cuba.

O que contradiz explicitamente (3). Ou considere (2) e (3). Elas implicam (por *modus tollens*)

(5) Não sei que tenho mãos.

O que contradiz explicitamente (1). A fim de evitar a contradição temos de abandonar pelo menos uma das afirmações (1)-(3). Mas qual?

2. Respostas ao puzzle

A rejeição de cada uma das afirmações (1)-(3) define, em linhas gerais, um tipo de resposta diferente ao nosso *puzzle*. Obviamente que seria uma tarefa difícil, neste artigo introdutório, exaurir todas as possíveis respostas; voltaremos nossa atenção então para as seguintes: (2.1) o ceticismo radical; (2.2) a resposta semântica; (2.3) o moorianismo; (2.4) a abordagem das alternativas relevantes; (2.5) a teoria rastreadora; (2.6) o neo-moorianismo, e (2.7) o contextualismo.

2.1. Rejeitando (1): o ceticismo radical

A solução cética é direta. Dado a plausibilidade de (2) e (3), nega-se (1). Temos então o *argumento cético*:

- 1 - Se sei que tenho mãos, então sei que não sou um cérebro numa cuba.
- 2 - Não sei que não sou um cérebro numa cuba.
- 3 - Logo, não sei que tenho mãos.

Esse argumento pode ser generalizado para cobrir todas as nossas crenças sobre o mundo externo, afinal a proposição <Tenho mãos> é apenas de uma das várias proposições sobre o mundo externo que consideramos saber. Façamos ME funcionar como uma variável para qualquer proposição acerca do mundo externo (e.g. <Tenho mãos>; <Ouro Preto fica em Minas Gerais>; <Há um computador defronte a mim>; etc.), e BIV representar a proposição <Sou um cérebro numa cuba>. O *Argumento Cético* generalizado pode ser assim então representado:

- 1 - Se sei que ME, então sei que não-BIV.
- 2 - Não sei que não-BIV.
- 3 - Logo, não sei que ME.

A solução cética não nos parece a mais atraente. Aceitar a conclusão de que nada sabemos sobre o mundo externo é um preço demasiado alto a se pagar a fim de se evitar o puzzle. Se quisermos, portanto, manter que sabemos algo sobre o mundo externo, temos de adotar outra estratégia e negar ou (2) ou (3).

É importante notar que uma resposta satisfatória ao argumento cético tem de satisfazer a pelo menos dois desideratos: (i) tem de apontar o erro do argumento cético, e (ii) explicar sua força – mais especificamente explicar por que é difícil rejeitar a hipótese cética. Uma resposta que satisfaça a (i) mas não a (ii) será defectiva no seguinte sentido: não conseguirá persuadir aqueles que de algum modo tendem a aceitar a conclusão cética.

Antes de prosseguir, permitam-me, porém, uma pequena digressão acerca do ceticismo.

Interlúdio I: o ceticismo filosófico e seu papel metodológico

Muitos tendem a pensar no ceticismo como uma maldição intelectual, e temem que a aceitação da conclusão cética nos leve a uma completa estagnação intelectual. Além do mais, tendem a pensar que aceitar a conclusão cética é reconhecer um certo tipo de fracasso de todo o nosso empreendimento cognitivo. Essa perspectiva um tanto dramática frente o argumento cético pode nos levar a adotar posturas bastante radicais. Uma delas é tentar bloquear o argumento cético desde o início sustentando que a adoção do ceticismo é autorrefutante. Uma vez que o cético precisa de premissas para sustentar a sua conclusão, há pelo menos algo que o cético tem saber, nomeadamente, as premissas de seu argumento. Assim, para sustentar que nada sabemos, o cético tem de saber algo; e, portanto, refuta a si próprio.

O argumento acima poderia ser efetivo contra um cético imprudente – se é que existe algum cético prudente! – que defendesse que não podemos saber sequer coisas como *tenho neste momento uma sensação de dor* ou *neste momento algo me parece vermelho*, etc. O cético do qual trataremos aqui sustenta apenas que nada podemos saber sobre o *mundo externo*. Se sabemos coisas introspectivamente, ou proposições da lógica e da matemática (e proposições conhecíveis *a priori* em geral) é algo que nosso cético não importa em conceder.

A outra postura, embora faça certa concessão ao cético, é radical no sentido de ser revisionista – o *idealismo*. O idealista concorda com o cético que não podemos saber que o mundo externo, constituído de objetos físicos, existe. Porém, não vê nisso uma ameaça ao nosso conhecimento da realidade objetiva, uma vez que pensa que a realidade é última instância mental, e não física. O cético se aproveita do hiato que o realismo abre entre a nossa mente e o mundo, da possibilidade de a causa das nossas representações mentais serem completamente diferentes das próprias representações. Ao negar o realismo, o idealista evita tal hiato. Assim, os estados mentais acerca da realidade são diretamente sobre a realidade, e não apenas meras representações dela. Para o idealista, então, podemos seguramente inferir que *p* a partir de *parece-me que p* (sem que haja a possibilidade de estarmos enganados). Poucos filósofos, porém, estão dispostos a pagar o preço do idealismo.

Há, contudo, uma atitude mais modesta frente ao ceticismo: podemos aceitar sua força e reconhecer seu papel metodológico para a epistemologia. Embora, talvez, o ceticismo seja extremamente difícil de se refutar, uma teoria do conhecimento (ou da justificação) adequada tem de lidar de alguma maneira com o desafio cético – seja explicando sua força, seja mostrando como não impugna nossas atribuições mais comuns de conhecimento, etc. Como dito na epígrafe que abre este texto, se o cético não existisse, precisaríamos inventá-

lo; ou de maneira mais pomposa, como bem assinalou Immanuel Kant em sua *Crítica da Razão Pura*:

o cético é o vigilante que conduz o raciocinador dogmático a uma saudável crítica do entendimento e da própria razão. [...] Em verdade, o método cético, em si mesmo, para as questões da razão, *não é satisfatório*, mas *preliminar*; serve para lhe despertar a cautela e indicar-lhe os meios sólidos que possam assegurar-lhe a legítima posse. (1781 A769)

Ao fazer isso, vários conceitos epistêmicos importantes podem ser clarificados, tais como *evidência, garantia, confiabilidade, acessibilidade, transparência, sorte epistêmica, transmissão, externo/interno, falibilidade* e assim por diante.

Desse ponto de vista, portanto, o cético está longe de ser o inimigo do epistemólogo; é, ao invés, seu *sparrring* – seu companheiro de treino! Dito isso, passemos às estratégias anticéticas.

2.2. Rejeitando (3): o externismo semântico

Hillary Putnam (1981) argumentou que a hipótese cética (do cérebro numa cuba) é inconsistente – se for verdadeira implica sua própria falsidade. O argumento depende de um pressuposto sobre como somos capazes de nos referir a objetos externos (ou igualmente à nossa capacidade de fazer representações do mundo externo). Daí termos uma resposta semântica ao ceticismo.

Suponhamos que eu seja um cérebro numa cuba. Assim, a proposição <Sou um cérebro numa cuba> é verdadeira. E se é verdadeira, então os termos “cérebro” e “cuba” se referem a algo. Mas se referem a quê? Supostamente a objetos extramentais, ao meu cérebro e à cuba na qual me encontro, respectivamente. É aqui que entram as idéias de Putnam sobre a referência. Uma palavra ou uma representação mental só se refere a algo na medida em que há uma conexão causal apropriada entre a palavra ou a representação e o objeto referido.⁵ Mas no caso em que fôssemos cérebros em cubas, os termos “cérebro” e “cuba”, por

5 - As idéias de Putnam sobre a referência e significado de nomes próprios e termos para categorias naturais encontram-se no seminal “The meaning of ‘meaning’” (1975). Putnam, ao lado de Saul Kripke (c.f. *Naming and Necessity* 1980), inaugurou aquilo a que se chama hoje *externismo semântico* – tanto na filosofia da linguagem quanto na filosofia da mente. Para uma exposição lúcida e informativa veja o *Philosophical Analysis in 20th Century*, de Scott Soames (2003).

exemplo, não estariam conectados causalmente (ou se têm qualquer conexão causal, essa não é apropriada) aos objetos externos; estariam, no máximo, conectados à *imagem do meu cérebro produzida pela simulação do supercomputador* e à *imagem da minha cuba produzida pela simulação*. Assim, se os termos “cérebro” e “cuba” se referem a algo, referem-se, no máximo, ao cérebro na imagem e à cuba da imagem. Ora, a hipótese do cérebro numa cuba diz respeito a cérebros e cubas reais, no mundo externo, não a cérebros e cubas virtuais. Assim, se é verdade que sou um cérebro numa cuba, quando digo que sou um cérebro numa cuba, estou dizendo algo diferente da hipótese original.

Simplificando o argumento: Se BIV, então “cérebro” e “cuba” referem-se a cérebro e cuba (no mundo real) respectivamente. Mas tais termos não se referem a cérebro e cuba. Logo, não é o caso que BIV.⁶

Suponhamos – deixando de lado possíveis dificuldades – que o argumento de Putnam seja sólido. Ele, contudo, não refuta a hipótese cética. Refuta, no máximo, a hipótese de que sempre fomos um cérebro numa cuba. Se alterarmos a hipótese – como apresentei já no início do texto – para a hipótese de que nossos cérebros foram removidos de nossos crânios, então nossos termos continuam se referindo ao mundo externo e não à imagem da simulação. O argumento cético ainda parece ficar de pé.

2.3. Rejeitando (3): a resposta mooriana

Outra maneira de se rejeitar (3) é através do seguinte argumento:

- 1 - Sei que tenho mãos.
- 2 - Se sei que tenho mãos, então sei que não-BIV.
- 3 - Logo, sei que não-BIV.

Esse tipo de argumento ficou conhecido como “mooriano” inspirado naquilo que G. E. Moore (1873-1958) chamou de uma prova da existência do mundo externo em “A proof of an external world” (1939).⁷ Embora a postura do Moore histórico frente ao ceticismo seja

⁶ - Não é consensual que assim formulado o argumento de Putnam seja capaz de sustentar a conclusão de que sei que não sou um BIV. Cf. Brueckner (1986; 2012) e Hickey (in IEP).

⁷ - O argumento acima não é exatamente o argumento de Moore. (Ele não faz qualquer alusão a cérebro numa cuba). Para ser

mais complexa do que o moorianismo,⁸ a ideia geral é de que dispomos de um argumento cogente cuja conclusão é “sei que não sou um cérebro numa cuba”.

Se prestarmos bem atenção, veremos que a segunda premissa do argumento mooriano é a primeira do argumento cético, de modo que para vencer o cético o mooriano só precisa argumentar a favor de sua primeira a premissa, a de que sei que tenho mãos. Qual o argumento? Talvez nem precisássemos de um; não é suficientemente óbvio que sabemos proposições de senso comum, *e.g.*, que tenho mãos, ou que há um computador em minha frente? Afinal, é uma questão de olhar e ver os objetos externos que estão à minha volta. Contudo, se eu fosse um BIV também “veria” mãos, computadores, e toda sorte de objetos (nesse caso, virtuais!). Portanto, simplesmente olhar e ver não parece não ser suficiente para que eu saiba que tenha mãos. Eu teria, além disso, de ser capaz de eliminar a hipótese cética como uma alternativa relevante na explicação daquilo vejo. Por conseguinte, tal estratégia mooriana não parece efetiva para resolver o puzzle cético.

O argumento mooriano é um caso típico de petição de princípio. Embora a conclusão – *sei que não sou um cérebro numa cuba* – não ocorra dentre as premissas do argumento, é pressuposta na defesa da primeira premissa, uma vez que eliminar a alternativa cética implica em saber que a hipótese cética é falsa. Isto é, o mooriano só pode sustentar que sabe que tem mãos se já souber que não é um cérebro numa cuba. E só pode sustentar que não é um cérebro numa cuba se já souber que tem mãos.

Outro problema suscitado pelo moorianismo é o que Stewart Cohen chamou de *conhecimento fácil*. Suponhamos (sem mais complicações) que sei perceptualmente que há uma mesa vermelha em minha frente. A proposição <Há uma mesa vermelha em minha frente> implica a proposição <A mesa em minha frente não é branca iluminada de modo que pareça vermelha>. (A primeira proposição implica a última porque não há qualquer circunstância possível em que seja verdade que a mesa em minha frente seja vermelha e que essa mesma mesa seja branca iluminada para parecer vermelha). Ora, se sei que a mesa é vermelha, e sei que se é vermelha não é branca iluminada para parecer vermelha, posso

mais preciso o argumento é algo mais ou menos como:

- 1 - Eis aqui uma mão.
- 2 - Eis aqui outra mão.
- C1 - Logo, há duas mãos.
- C2 - Logo, há objetos externos.

Certamente que podemos, sem problema algum, prefixar o operador de conhecimento “Sei que” às premissas e às conclusões do argumento acima.

Não é exatamente claro que a estratégia de Moore seja contra o cético; a sua tentativa de provar que o mundo externo existe pode ser entendida como uma refutação do idealismo (cf. BALDWIN 1993). Contudo, uma vez que a motivação principal idealista é evitar o ceticismo – rejeitando o mundo externo – uma prova do mundo externo que refute o idealista também refuta o cético.

8 - Talvez Moore não tenha sido um mooriano! Veja BAUMANN (2008).

inferir que a mesa não é branca iluminada para parecer vermelha e, através disso, saber que a mesa não é branca iluminada para parecer vermelha. (É exatamente o mesmo tipo de raciocínio usado pelo mooriano). Contudo, o tipo de evidência adequada para tornar conhecimento a minha crença de que há uma mesa vermelha em minha frente é diferente do tipo de evidência adequada para a crença de que a mesa não é branca iluminada para parecer vermelha. A mesa parecer vermelha é geralmente considerado como boa evidência para a crença de que a mesa é vermelha; contudo, não parece ser uma boa evidência para a crença de que a mesa não é branca iluminada para parecer vermelha. Se o mooriano estiver certo, então parece que podemos obter conhecimento de maneira demasiado fácil de certas proposições que são implicadas por proposições de senso comum – como a proposição de que tenho mãos ou a de que há uma mesa vermelha na minha frente. Intuitivamente, porém, não parece ser esse o caso.

Ademais, o mooriano não é capaz de explicar a força do argumento cético. Supor simplesmente que sabemos proposições do senso comum não nos dá qualquer pista de por que parece plausível dizer que não sabemos as negações das hipóteses céticas.⁹

2.4. Rejeitando (2): alternativas relevantes

Vimos na seção anterior que uma das dificuldades de sustentar que tenho mãos é que não é fácil afastar a alternativa cética. Para tornamos isso mais claro, podemos pensar na seguinte exigência para o conhecimento:

ALT:

Se S sabe que p , então, para qualquer q relevante incompatível com p , S é capaz de eliminar q .

Por exemplo, suponha que Deckard, de passagem por Ouro Preto, veja Joaquim na praça e forme a crença verdadeira de que ali há um homem. Geralmente atribuiríamos conhecimento a Deckard. Mas desconhecido a ele, Ouro Preto se tornara uma colônia

⁹ - Duncan Pritchard (2007, pp. 70-75) oferece um diagnóstico conciso das principais falhas do moorianismo (algumas delas supracitadas). Ele aponta seis dificuldades centrais para o argumento mooriano: (i) petição de princípio contra o argumento cético; (ii) o impasse gerado pela força da premissa 2 do argumento cético; (iii) a inadequação conversacional do uso de "Eu sei que p "; (iv) falta de uma teoria de fundo que explique como podemos saber que não-BIV; (v) não explica a força do argumento cético, e (vi) não é capaz de lidar com a subdeterminação.

de replicantes Nexus-6 – que são visual e comportamentalmente indistinguíveis de seres humanos. Joaquim, por sorte, é o último humano de Ouro Preto. Já agora, hesitamos bastante em atribuir conhecimento a Deckard – muito embora a sua crença seja verdadeira e formada de maneira adequada. A razão pela qual hesitamos em atribuir conhecimento a Deckard é que ele não é capaz de eliminar uma alternativa relevante, a de que Joaquim, ao invés de ser homem, é um replicante. Essa alternativa é relevante porque tivesse Deckard olhado para o Padre Rolim, e não para Joaquim, teria acreditado falsamente que havia na praça uma pessoa.

O cético parece se aproveitar de ALT. Uma vez que não somos nunca capazes de afastar a hipótese BIV – que explica igualmente os meus dados dos sentidos – não podemos saber qualquer coisa que seja sobre o mundo externo. Mas será que BIV é uma alternativa relevante?

Consideremos primeiro o seguinte caso, apresentado por Fred Dretske (1970). Levo meu filho ao zoológico e lhe aponto um animal que se parece uma zebra – tem porte de equino, é listrado, etc. – dentro de uma jaula em que há uma placa com a indicação “zebra”. É verdade que o animal é uma zebra, e meu filho forma a crença verdadeira, com base em evidências adequadas, de que o animal é uma zebra. Contudo, há a possibilidade de aquele animal ser, ao invés de uma zebra, uma mula pintada para parecer uma zebra (assim como no caso da mesa apresentado na objeção do conhecimento fácil). Talvez o dono do zoológico não tenha tido dinheiro suficiente para comprar uma zebra, e para não perder freguesia, disfarçou a mula de zebra. A proposição <o animal é uma mula disfarçada> é uma alternativa à proposição <o animal é uma zebra>, mas não é relevante, pensa Dretske. As circunstâncias possíveis – ou mundos possíveis – em que o animal fosse uma mula disfarçada seriam distantes o bastante da situação atual de modo que a possibilidade de a proposição <o animal é uma mula disfarçada> ser verdadeira não ameaça o estatuto epistêmico da minha crença, no mundo atual, de que o animal é uma zebra.

A hipótese cética é como a hipótese de o animal ser uma mula disfarçada: as circunstâncias possíveis em que é verdadeira são distantes o suficiente da circunstância atual (ou efetiva) que a sua mera possibilidade não afeta o estatuto epistêmico da minha crença de que tenho mãos, ou que há um computador em minha frente. Portanto, a hipótese cética não é uma alternativa relevante à minha crença de que tenho mãos.

Note, agora, que a proposição <o animal é uma zebra> implica <o animal não é uma mula disfarçada>, uma vez que não há qualquer circunstância possível em que a primeira seja verdadeira e a última falsa. O leitor pode então estar perplexo: uma vez que sei a implicação, sei a antecedente, basta que raciocine e conclua que sei que o animal não é

uma mula disfarçada (o que mostra que sou capaz de afastar a alternativa). Porém, Dretske argumenta que esse é um caso em que não posso saber a implicação, pelo menos não desse modo. As evidências de que disponho são adequadas para a crença de que o animal é uma zebra – se o animal fosse um hipopótamo, por exemplo, não me pareceria uma zebra e eu não formaria a crença de que o animal é uma zebra; contudo, não são evidências adequadas para a crença de que o animal não é uma mula disfarçada – uma vez que ainda que houvesse uma mula disfarçada na jaula, eu ainda continuaria acreditando que o animal é uma zebra. (Lembre-se do caso de Deckard – ele não sabe que Joaquim é um humano pois tivesse Joaquim sido um Nexus 6, ele ainda continuaria acreditando que Joaquim fosse humano). Assim, sei que o animal é uma zebra, mas não sei que não é uma mula disfarçada.

O mesmo vale para a condicional do cético: sei que tenho mãos, mas não sei que não sou um cérebro numa cuba. Isso parece mostrar que a proposição 2 do puzzle cético é falsa. Essa alternativa mantém a força da proposição 3 – <Não sei que não sou um cérebro numa cuba> –, mas tira-lhe a importância: não é uma alternativa relevante às nossas crenças sobre o mundo externo.

2.5. Rejeitando (2): a teoria rastreadora

Uma abordagem bastante próxima à anterior se deve a Robert Nozick (1981). Nozick, e também Dretske (1971), argumentou que o conhecimento possui uma componente contrafactual – *i.e.* uma crença verdadeira conta como conhecimento quando certas relações entre o modo como as coisas são e como poderiam ter sido são satisfeitas. Na seção anterior, ao explicar porque certas alternativas não são relevantes usamos o termo “formaria”, no subjuntivo – *se tais e tais coisas fossem de tal modo, eu formaria a crença tal*. A ideia de Nozick é que se sei uma proposição p qualquer, então se não fosse o caso que p eu não *formaria* a crença de que p .¹⁰ Se uma crença conta como conhecimento, ela tem de ser sensível (ou sensitiva) à verdade num conjunto determinado de circunstâncias.¹¹

*Sensitividade: se p não fosse o caso, então S não formaria a crença de que p .*¹²

10 - Ele também adiciona uma outra condição – se p fosse o caso, então eu formaria a crença de que p – a que ele chama de condição de aderência. Podemos, contudo, deixá-la de lado aqui.

11 - A versão da teoria rastreadora sustentada por Dretske (1971), cuja teoria de Nozick é uma modificação, é conhecida como “abordagem das razões conclusivas”. S sabe que p se e somente se tem uma razão conclusiva R para p . R, por sua vez, é uma razão conclusiva para p se e somente se não fosse p o caso, R não teria sido o caso.

12 - Para uma formulação mais precisa e que evite contraexemplos temos de fazer referência ao método ou processo pelo

Para tornar mais precisa essa noção temos de introduzir o jargão dos mundos possíveis. Ao conjunto de circunstâncias nas quais as coisas poderiam ter sido de determinado modo chamaremos “mundos possíveis”. Ao modo como as coisas efetivamente são, *i.e.* a circunstância efetiva ou atual, chamamos “mundo atual” (ou “efetivo”). Por exemplo, no mundo atual a proposição <Osama Bin Laden é um terrorista> é verdadeira, mas há vários mundos possíveis nos quais é falsa. Há mundos possíveis nos quais ao invés de despender suas forças para o terror, ele ajuda a eliminar a pobreza no mundo.¹³

Diremos agora que uma condicional contrafactual do tipo *se fosse o caso que p, então seria o caso que q* significa aproximadamente *nos mundos possíveis mais próximos ao atual em que p é verdadeira, q também é verdadeira*. Assim, *se não fosse o caso que p então eu não acreditaria que p* significa *nos mundos possíveis mais próximos ao mundo atual em que p é falsa eu não formo a crença de que p*. Há muitos mundos possíveis em que *p* é falsa e ainda assim formo a crença de que *p*; contudo, só os mundos mais próximos é que contam para a avaliação da condicional contrafactual: se nesses mundos próximos *p* é falsa e não formo a crença de que *p*, então a condicional é verdadeira; se, por outro lado, *p* é falsa e ainda assim formo a crença de que *p*, então a condicional é falsa.¹⁴

A noção de sensibilidade torna-se agora um pouco mais clara: sei uma proposição *p* quando nos mundos possíveis mais próximos ao atual em que *p* é falsa (ou não é o caso) não formo a crença de que *p*. Quando sabemos que *p* a nossa crença de que *p* “rastrea” a verdade nesse conjunto de mundos possíveis. Fica mais clara também a noção de alternativa relevante: a extensão das alternativas relevantes é delimitada ao conjunto de mundos possíveis próximos.

Estamos prontos agora para entender a resposta de Nozick ao ceticismo. Primeiro usamos

qual o sujeito formou sua crença. Assim, a sensibilidade deveria ser lida do seguinte modo, com “M” representando o método ou processo de formação de crença: *se não fosse o caso que p, então S não acreditaria, via M, que p*. Para facilitar deixaremos implícita a menção ao método.

13 - De maneira um pouco mais precisa, um mundo possível é um conjunto maximal consistente de proposições. Isto é, para qualquer proposição, ou essa proposição pertence ao conjunto ou a sua negação pertence. Isso nos garante que tenhamos uma descrição total do mundo. O mundo atual é também um mundo possível. A princípio não há restrições à quantidade de proposições diferindo em valor de verdade do mundo atual para que um conjunto de proposições seja considerado um mundo possível não-atual. Por exemplo, uma descrição total do mundo que contenha todas as proposições que são verdadeiras no mundo atual menos uma, é um mundo possível não-atual; e assim por diante.

14 - Não é exatamente fácil determinar que conjunto de mundos possíveis conta como o *mais próximo* ao atual. Intuitivamente, um mundo possível será tanto mais próximo ao atual quanto menos proposições diferirem em valor de verdade. Por exemplo, suponhamos que o conjunto @ possui apenas três proposições verdadeiras, {*p, q, r*}, e que em *w1* *p* é falsa, e em *w2* *p* e *q* são falsas. Nesse caso, *w1* é mais próximo a @ do que *w2*. Seguiremos a sugestão de Dretske (1971) (embora não sem dificuldades) de que os mundos próximos relevantes para a atribuição de conhecimento são aqueles em que apenas a proposição alvo (e todas aquelas proposições que dependem dela lógicamente e causalmente) tem seu valor de verdade alterado; tudo o mais mantemos inalterado. Para dificuldades sobre essa proposta veja o exemplo do termômetro em GOLDMAN (1986, p. 45).

a sensibilidade para explicar o nosso conhecimento de proposições de senso comum, *e.g.*, a de que tenho mãos. Vejamos. Se sei que tenho mãos, então se não fosse o caso que tenho mãos, eu não acreditaria que tenho mãos. No jargão dos mundos possíveis: nos mundos possíveis mais próximos em que não tenho mãos não formo a crença de que tenho mãos. Os mundos mais próximos em que não tenho mãos são, por exemplo, os mundos em que perdi as mãos num acidente, ou os mundos em que nasci sem mãos, etc. Nesses mundos, quando olho para os cotocos nos meus braços, digamos, não formo a crença (falsa) de que tenho mãos. Portanto, de acordo com a sensibilidade, sei que tenho mãos. Ou considere a minha crença verdadeira de que há um computador em minha frente neste exato instante. Nos mundos mais próximos nos quais não é o caso que há um computador em minha frente – *e.g.* nos mundos em que estou redigindo este texto à mão – quando olho para o texto, não formo a crença de que há um computador em minha frente. Portanto, de acordo com a sensibilidade, sei que há um computador em minha frente. O conhecimento é, de acordo com tal perspectiva, crença verdadeira sensitiva.

O segundo passo consiste em mostrar que a minha crença de que não sou um cérebro numa cuba não é sensitiva. Ou seja, temos de mostrar que nos mundos possíveis mais próximos em que a proposição <Sou um BIV> é verdadeira, ainda assim formo a crença de que não sou um BIV. E isso não é difícil de ver, pois faz parte da hipótese cética de que sou um cérebro numa cuba que não sou capaz de distinguir em primeira pessoa se minhas experiências sensoriais foram causadas pelo mundo externo ou pelos impulsos enviados do supercomputador. Assim, naqueles mundos em que sou um cérebro numa cuba – e que estou massivamente enganado sobre o mundo externo – ainda assim formo a crença de que não sou um cérebro numa cuba, pois infiro a partir crença falsa de que tenho mãos que não sou um BIV. De acordo com a sensibilidade, portanto, não sei que não sou um cérebro numa cuba.

O terceiro passo é simples: é só conjuntarmos a conclusão do primeiro passo – *sei que tenho mãos* – com a conclusão do segundo – *não sei que não sou um cérebro numa cuba*. Ora, essa conjunção é equivalente a *não é o caso que se sei que tenho mãos, então sei que não sou um cérebro numa cuba*. Ou seja, o resultado de se aplicar o teste da sensibilidade a ambas as crenças é a negação da condicional que constitui a proposição (2) do puzzle cético.

Em suma, a sensibilidade concede ao cético a força de sua hipótese, mas nega que a nossa incapacidade de excluí-la nos prive de obter conhecimento.

2.5.1. A conjunção abominável: rejeição do princípio de fechamento

Embora promissora, a teoria rastreadora (juntamente da teoria das alternativas relevantes de Dretske) enfrenta um desafio considerável. Uma vez que a proposição <Tenho mãos> implica a proposição <Não sou um BIV> – *i.e.* se tenho mãos, então não sou um cérebro numa cuba – soa bastante estranho dizer que sei que tenho mãos, mas que não sei que não sou um cérebro numa cuba. É como se eu dissesse “Sei que estou escrevendo agora, (e se estou escrevendo agora, é verdade que estou vivo), mas não sei que estou vivo”. Keith DeRose (1995) chamou a isso “conjunção abominável”.

O que torna essa conjunção abominável é que o conhecimento parece ser fechado sob implicação conhecida. Isto é, se sei uma proposição qualquer, sei que essa proposição implica um outra proposição, então também sei essa última. Assim, o seguinte princípio de fechamento epistêmico parece verdadeiro (no qual K é o operador de conhecimento e significa “S sabe que”):

CK:

$$[K\alpha \ \& \ K(\alpha \rightarrow \beta)] \rightarrow K\beta^{15}$$

Não é difícil ver como CK sustenta a proposição (2) do puzzle. É plausível pensar que a proposição <Se tenho mãos, então não sou um cérebro numa cuba>, que vamos representar por $(p \rightarrow \neg q)$, é uma verdade conceitual. Basta pensarmos sobre o conceito de *mundo externo* e na hipótese cética para percebermos que não pode ser verdade que ao mesmo tempo que tenho mãos também sou um BIV, de modo que $K(p \rightarrow \neg q)$. Assim, sob a suposição de que sei que tenho mãos, Kp , e dado que CK é verdadeiro, se sei que tenho mãos, então sei que não sou um cérebro numa cuba, $Kp \rightarrow K\neg q$. Esquemáticamente:

1 - Kp	(Suposição)
2 - $K(p \rightarrow \neg q)$	(Verdade conceitual)

15 - Assim formulado o princípio enfrenta contraexemplos óbvios: posso saber que p , saber que p implica q , e, contudo, não acreditar que q , de modo que também não posso saber que q . Isso pode ocorrer pois posso não ter feito uma inferência e concluído que q . Ou posso simplesmente ter formado a crença em q aleatoriamente. Assim o princípio tem de fazer referência a S acreditar em q a partir de uma inferência apropriada de p e $p \rightarrow q$. Mas isso ainda não é suficiente. Poderia ser o caso de eu ter formado a crença em q de maneira apropriada, mas no momento posterior ela ter sido apagada da minha memória. Assim, temos também de fazer referência à retenção da crença. (Para uma breve discussão sobre a formulação do fechamento veja LUPER 2016). Para os nossos propósitos, contudo, o princípio simplificado CK é suficiente.

3 - $[Kp \ \& \ K(p \rightarrow \neg q)] \rightarrow K\neg q$	(Instância de CK)
4 - $Kp \ \& \ K(p \rightarrow \neg q)$	1, 2 I&
5 - $K\neg q$	3, 4 MP
6 - $Kp \rightarrow K\neg q$	1- 6 I \rightarrow

Se pensamos que $Kp \rightarrow K\neg q$ é falsa, só nos resta negar o passo 2 ou o 3, uma vez que a falsidade de 6 implica $Kp \ \& \ \neg K\neg q$. Mas 2 é bastante implausível de se negar, de modo que temos de negar o passo 3. Ora, negar uma instância de um princípio geral é mostrar que esse princípio é falso. Portanto, temos de negar CK.

O problema de se rejeitar o princípio de fechamento é que ele parece ser indispensável para explicar como podemos saber algo inferencialmente. Ou seja, o fechamento é pelo menos uma condição necessária ao conhecimento inferencial. Por exemplo, é parte da explicação do conhecimento inferencial que o sujeito tem de saber as premissas de sua inferência e que sua garantia tem de ser, de algum modo, transferida para a conclusão. Tem de haver uma contraparte epistêmica da preservação da verdade; o princípio de fechamento parece oferecer tal contraparte.

Se, por um lado, a rejeição do fechamento afasta o desafio cético, por outro nos deixa sem uma explicação de como podemos saber algo inferencialmente.

Nozick está ciente de que a sua teoria implica a falsidade do princípio de fechamento. Contudo, pensa que essa não seja uma consequência ruim. Na verdade, ele pensa que a sua teoria contrafactual explica exatamente porque o fechamento falha. Seu argumento, porém, é falacioso. Ele argumenta que porque uma das condições necessárias ao conhecimento, a sensibilidade, não é fechada sob implicação conhecida, o conhecimento também não é. Ora, a crença, uma condição necessária ao conhecimento, claramente não é fechada sob implicação conhecida, e nem por isso dizemos que se segue disso que o conhecimento não é fechado sob a implicação conhecida. Do fato de uma das condições necessárias ao conhecimento não ser fechada sob implicação conhecida não se segue o conhecimento não seja. Pode ser que uma ou mais condições juntas é que sejam responsáveis pelo fechamento. Argumentar como Nozick é cometer a falácia de composição (cf. WARFIELD 2004). A despeito disso, a teoria de Nozick é robusta o suficiente para apoiar a negação do fechamento. Por exemplo, o fato dela lidar de maneira bem sucedida com vários casos à la Gettier. Além do mais, há quem esteja disposto abandonar o fechamento caso seja esse o preço a pagar para resolver problemas epistemológicos importantes como a natureza do conhecimento e

o ceticismo (cf. BECKER 2007, pp. 113-28; ADAMS e CLARKE 2005, p. 207, nota 2).¹⁶

Saul Kripke (2011), num influente artigo sobre a epistemologia de Nozick, apresentou um problema mais agudo à rejeição do fechamento. Suponha que você esteja dirigindo por uma região peculiar: seus moradores têm a tradição de erigir celeiros *fake* (de fachada), todos da cor azul. Dos 100 celeiros erigidos, 99 são *fake*, os azuis, e apenas 1 deles real, esse pintado de vermelho. De acordo com a sensibilidade, você, ao olhar para o celeiro vermelho e formar a crença de que está defronte a um celeiro vermelho, sabe que está de frente a um celeiro vermelho, pois a sua crença rastreia a verdade: nos mundos possíveis mais próximos em que você não está defronte a um celeiro vermelho, mas antes defronte a um azul, você não forma a crença de que está diante de um celeiro vermelho. Note também que a proposição <Há um celeiro vermelho defronte a mim> implica a proposição <Há um celeiro defronte a mim>. Contudo, de acordo com a sensibilidade, não sei que há um celeiro defronte a mim, uma vez que nos mundos próximos em que não há um celeiro defronte a mim, ainda acredito que há. Essa negação do fechamento parece ser muito mais estranha (e abominável) do que a de que sei que tenho mãos e não sei que não sou um BIV.

Há várias outras dificuldades – às quais não precisamos nos deter – que a teoria rastreadora enfrenta. Algumas técnicas, como por exemplo, o modo de se determinar quais mundos possíveis contam como próximos e quais não contam. Outras, mais graves, dizem respeito ao alcance: há argumentos tanto a favor de que a sensibilidade é demasiado estrita, quanto a favor de que é demasiado permissiva (para uma apresentação clara e sucinta, além de discussão, dessas objeções veja BECKER 2007).

Fred Dretske (1970) tentou explicar a falha do fechamento argumentando que operadores epistêmicos – e.g. “S sabe que”, “S tem evidência para”, “S está justificado”, etc. – são o que ele chamou de *semipenetrantes*. De acordo com ele, um operador é *penetrante* quando é fechado sob implicação conhecida (ou justificada, ou garantida, etc.). Uma vez que o conhecimento viola o fechamento (vide os exemplos da mula disfarçada na seção 2.4, e do cérebro numa cuba na seção 2.5), o operador de conhecimento “S sabe que” não é penetrante. Um operador é *semipenetrante* quando penetra em alguns casos mas não em outros. Por exemplo: $(Kp \ \& \ Kq) \rightarrow Kp$. Se S sabe essas duas proposições, então é verdade que sabe apenas uma delas. Nesse caso o operador “S sabe que” – o operador K – penetra na proposição implicada. Mas no caso da mula disfarçada, por exemplo, K parece não penetrar. Dretske não tem um argumento direto a favor disso; ao invés ele argumenta que há outros

16 - Peter Baumann (2012) argumenta que, sem que o próprio Nozick tenha se apercebido, é possível sustentar uma versão mais sofisticada do fechamento que esteja de acordo com a sensibilidade. Sherrilyn Roush (2006) tenta restabelecer a teoria rastreadora de Nozick, mantendo o fechamento, mas recorrendo não às condicionais contrafactuais, e sim a um enquadramento bayesiano.

operadores epistêmicos, e.g., o operador de explicação, que são também semipenetrantes.

Peter Klein (1995) aponta uma ambigüidade na argumentação de Dretske. Ao invés de ter refutado o princípio de fechamento, o que ele refutou foi, na verdade, o princípio de transmissão de evidência. O contraexemplo da zebra parece bem sucedido porque considera que o animal parecer uma zebra, embora seja boa evidência para a crença de que o animal é uma zebra, não é uma boa evidência para a crença de que o animal não é uma mula disfarçada. Ou seja, o status epistêmico conferido pela evidência na primeira crença não é transmitido para a segunda. Contudo, Klein nos atenta para uma leitura mais adequada do princípio de fechamento: não é a evidência a favor da primeira crença que tem de ser adequada para a segunda, mas antes é a própria crença (garantida, justificada, etc.) que irá funcionar como evidência. Isto é, a antecedente (quando sustentada) de uma instância do princípio de fechamento contará como evidência para a sua conseqüente.

Em sua última contribuição à discussão sobre o princípio de fechamento, Drestke (2005) tenta explicar a falha do fechamento recorrendo à ideia das *implicações de grande porte* [*heavyweight implications*]. As implicações de grande porte são algo como o seguinte: *se p* (em que *p* representa uma proposição do senso comum) *então não-BIV*, *se p então há um mundo externo*, *se p então há objetos físicos*, etc. Dretske argumenta que qualquer inferência que envolva implicações de grande porte violará o princípio de fechamento e, portanto, falhará em fornecer conhecimento da conclusão ao sujeito.

Interlúdio II: regresso ad infinitum, subdeterminação e ceticismo

É importante notar que nem toda formulação do argumento cético faz uso do princípio de fechamento epistêmico. Por exemplo, o argumento pirrônico não depende diretamente do fechamento. Vejamos:

A - Se sei uma proposição qualquer P_1 , então sei outra proposição P_2 que implica P_1 .

B - Se sei P_2 , então sei P_3 ...

C - Se sei P_n , então sei P_{n+1} e (ou seja, *ad infinitum*)

D - Assim, sempre haverá uma proposição anterior P_{n+1} que não sei.

E - Por conseguinte (por sucessivos *modus tollens*), não saberei $P_{n-1}, \dots, P_n, \dots, P_2, P_1$.

O argumento pirrônico estabelece a conclusão cética através do regresso *ad infinitum* de proposições. Uma vez que a exigência para saber uma proposição é saber outra, anterior a ela, sempre haverá uma proposição que será preciso saber, de modo que nunca atingiremos uma primeira proposição. Como se vê, esse argumento não faz menção ou mesmo pressupõe o fechamento.

Outro argumento cético que dispensa o fechamento é o argumento da subdeterminação. Duas hipóteses inconsistentes entre si, h e h' , e consistentes com um dado qualquer e são subdeterminadas se, e somente se, têm exatamente as mesmas conseqüências observacionais o . Assim, a observação de o não é suficiente para decidir entre h e h' ; por conseguinte, temos de suspender o juízo quanto à verdade de ambas. Vejamos como a subdeterminação gera o argumento cético:

A - Se a hipótese do mundo externo é verdadeira, então explica a minha experiência visual de ver minha mão.

B - Se a hipótese do cérebro numa cuba é verdadeira, então explica a minha experiência visual de ver minha mão.

C - A minha experiência visual não é suficiente para decidir entre as hipóteses.

D - Logo, tenho de suspender meu juízo quanto à origem das minhas experiências visuais da minha mão.

Novamente o argumento não faz menção ou pressupõe o fechamento (cf. BRUECKNER 1994, 2005; PRITCHARD 2005 e BRIESEN 2010 para uma discussão sobre a subdeterminação e o fechamento).

Dado o nosso puzzle cético inicial ter sido formulado com a pressuposição do fechamento, não daremos atenção a essas outras versões do argumento cético. Isso não impede, porém, que as alternativas aqui apresentadas sejam capazes de lidar com tais argumentos.

2.6. De volta à (3): o neo-moorianismo

Vimos que a teoria rastreadora enfrenta problemas ao rejeitar o fechamento. Mas daí não se segue que tenhamos de rejeitar qualquer abordagem contrafactual (ou modal) ao conhecimento. Uma abordagem modal alternativa, que usa a contrapositiva¹⁷ da sensibilidade, está disponível:

*Segurança: Se S acreditasse que p , então p seria o caso.*¹⁸

No jargão dos mundos possíveis, nos mundos possíveis mais próximos em que S acredita que p , p é verdadeira.

A segurança se sai igualmente bem nos vereditos dos casos à la Gettier, porém tem a suposta vantagem de manter o fechamento. Além disso, parece dar conta do conhecimento indutivo, coisa que a sensibilidade parece não fazer (SOSA 1999).

A segurança não viola o fechamento pois permite que possamos saber a negação da hipótese cética a partir de conhecimento do senso comum. Assim, o argumento mooriano é restabelecido – daí termos uma resposta chamada “neo-moorianismo” (PRITCHARD 2002, 2007). Para vermos que a segurança funciona, basta aplicar o teste contrafactual, assim como fizemos com a sensibilidade. Nos mundos mais próximos em que acredito que tenho mãos é também verdade que tenho mãos. É claro que há mundos em que acredito que tenho mãos e é falso que tenho mãos – *e.g.* no mundo BIV – mas esses mundos não contam como próximos. Assim, de acordo com a segurança, sei que tenho mãos. A condicional, *se tenho mãos, então não sou um BIV* – também conta como conhecimento, uma vez nos mundos próximos em que acredito nela (refletindo sobre o conceito de mundo externo) ela continua sendo verdadeira. Por último, a crença de que não sou um BIV. Nos mundos próximos em que acredito que não sou um BIV – *e.g.* o mundo em que não tenho mãos – é também verdade que não sou um BIV. Novamente, é claro que há mundos em que acredito que não sou um BIV e sou de fato um BIV; mas esses mundos são distantes o suficiente para não contarem como mundos próximos. Assim, sei que não sou um BIV. E não há qualquer violação do fechamento: sei que tenho mãos, sei que se tenho mãos não sou um BIV, e sei que não sou um BIV.

A segurança e a sensibilidade, à partida, dão os mesmos vereditos sobre os casos comuns de conhecimento. A diferença crucial entre elas fica clara quando consideramos as negações

¹⁷ - A contrapositiva de uma contrafactual, diferentemente de uma condicional material, não é equivalente. Assim, uma condicional sensitiva não tem as mesmas condições de verdade que uma condicional segura. A extensão do conjunto de mundos próximos em que avaliamos a condicional da sensibilidade não coincide com a extensão de mundos próximos em que avaliamos a condicional da segurança.

¹⁸ - Assim como no caso da sensibilidade, devemos também manter implícita a referência ao método ou processo de formação de crença.

das hipóteses céticas. Como vimos, de acordo com a sensibilidade não sabemos as negações das hipóteses céticas; mas de acordo com a segurança podemos saber. Isso ocorre porque a sensibilidade, ao levar em conta os mundos em que a proposição alvo é falsa, considera como próximos um conjunto de mundos que a segurança considera distantes o suficiente.

Embora a segurança pareça preferível à sensibilidade, não está livre de problemas; há contraexemplos à necessidade da segurança como condição ao conhecimento (cf. COMESAÑA 2005). E também não é claro que seja capaz de lidar com o contraexemplo da violação do fechamento levantado por Kripke (seção 2.5.1 acima).

2.7. Rejeitando o invariantismo: o contextualismo

Até agora todas as tentativas de solução ao puzzle cético, embora diversas, têm algo relevante em comum – pressupõem que os padrões a serem satisfeitos para que se atribua conhecimento a um sujeito são invariantes. O contextualismo, por outro lado, sustenta que os padrões para atribuição do conhecimento variam de acordo com o contexto. Uma maneira pela qual poderíamos chegar naturalmente ao contextualismo é através das alternativas relevantes. A ideia da abordagem das alternativas relevantes é que para saber que p não precisamos excluir todas as alternativas a p , apenas as alternativas relevantes. E o que determina quais são as alternativas relevantes e quais não são é o contexto. Por exemplo, suponha que você está numa festa à fantasia e vê alguém fantasiado de Gandolf. É natural atribuímos conhecimento a você quando você aponta para a pessoa e diz “Eis ali o meu amigo”. Mas suponha que haja mais três Gandolf’s ali; todos fantasiados perfeitamente e indistinguíveis uns dos outros à primeira vista. Já nesse caso não lhe atribuiríamos conhecimento, pois você não é capaz de excluir (não visualmente) as alternativas relevantes à proposição de que ali está o seu amigo. Se fosse o seu desafeto ao invés de seu amigo você ainda acreditaria que é o seu amigo. Ou seja, no primeiro contexto você sabe que p , ao passo que no segundo contexto você não sabe essa mesma proposição.¹⁹ Alguns epistemólogos sustentam que nossas atribuições de conhecimento – e, por conseguinte, se o sujeito sabe ou não – são sensíveis ao contexto, como o exemplo recém mencionado.

O contextualismo epistêmico é visto geralmente como uma tese semântica, *i.e.*, uma tese

¹⁹ - Não há a teoria contextualista. Na verdade, há várias teorias diferentes. O contextualismo motivado pela abordagem de Dretske (e.g. HELLER 1999) tende a negar o fechamento epistêmico, enquanto que Steven Cohen (1988) defende uma abordagem das alternativas relevantes que mantém o fechamento. Um tipo diferente de contextualismo, que também mantém o fechamento, é o baseado na sensibilidade (DEROSE 1995). Por motivos de espaço, apresentarei apenas a abordagem baseada na sensibilidade.

sobre as condições de verdade de proposições do tipo *S sabe que p*. A ideia que é o termo “sabe” funciona como as expressões indexicais – “eu”, “ele”, “aqui”, etc. –, casos paradigmáticos de expressões cujo significado é sensível ao contexto. Por exemplo, a frase *Eu estou em Ouro Preto*, proferida por mim neste instante, é verdadeira, pois expressa a proposição <L. H. está em Ouro Preto>; mas falsa se proferida neste instante por Donald Trump, pois expressa a proposição diferente <D. T. está em Ouro Preto>. De modo análogo a frase *Mick Jagger sabe que o Brasil foi derrotado pela Alemanha*, mediante a contextos epistêmicos diferentes, pode expressar proposições diferentes.

Os invariantistas consideram que o predicado “sabe que” expressa uma relação binária entre um sujeito e uma proposição. Já os contextualistas pensam que “sabe que” expressa uma relação triádica entre o sujeito, a proposição e o contexto. É por isso que os contextualistas pensam que *S sabe que p* pode expressar proposições diferentes: o terceiro elemento da relação pode variar. A frase “S sabe que p” não expressará uma proposição a menos que o contexto esteja nela indexado.

Mas como o contextualismo pode nos ajudar a responder ao desafio cético? Keith DeRose (1995) argumenta que nos contextos comuns, cotidianos, em que as exigências ao conhecimento são mais brandas, sei que tenho mãos e sei também que não sou um cérebro numa cuba; ao passo que nos contextos céticos, cujas exigências são mais elevadas, não sei que tenho mãos, pois não sei que não sou um cérebro numa cuba. DeRose recorre à sensibilidade para explicar a força da posição epistêmica que um sujeito tem para satisfazer as exigências de um determinado contexto – e, por conseguinte, tornar verdadeira uma atribuição de conhecimento a esse sujeito. Nos contextos cotidianos, em que preocupações céticas não estão em jogo, o conjunto de mundos possíveis próximos não inclui qualquer mundo em que a proposição <Eu sou um BIV> é verdadeira – os mundos-BIV são distantes do mundo atual. Porém, em contextos céticos, em que somos desafiados a mostrar que sabemos que não estamos num mundo-BIV, a esfera de mundos possíveis próximas é ampliada de modo a incluir mundos-BIV. Assim, dada a sensibilidade, nos contextos cotidianos a minha crença de que tenho mãos rastreia a verdade dessa proposição nos mundos próximos – mundos em que não sou um BIV; por outro lado, nos contextos céticos, há pelo menos um mundo em que a minha crença de que não tenho mãos não rastreia a verdade da proposição <Tenho mãos>, pois no mundo-BIV mais próximo em que acredito que tenho mãos, não é o caso que tenho mãos.

Em outras palavras, o argumento mooriano pode ser aplicado com sucesso no contexto cotidiano, mas não no contexto cético. Contudo, o que nos importa é o contexto cotidiano, não o contexto cético. Assim, o contextualista diz que não há de fato inconsistência entre

as proposições (1)-(3). Em contextos mais brandos, (1) é verdadeira e (3) é falsa; ao passo que em contextos mais elevados – e.g. num cenário cético – (3) é verdadeira e (1) falsa.

A vantagem dessa resposta é que, ao contrário das abordagens rastreadoras e das alternativas relevantes, o princípio do fechamento epistêmico é mantido. É por isso que o argumento mooriano pode ser aplicado no contexto cotidiano, mas não no contexto cético. Há, porém, uma restrição: o fechamento é válido apenas dentro do contexto, não entre contextos – i.e., vale apenas *intra* e não *inter*-contextos. Assim, não é permitido que façamos a seguinte inferência: sei que tenho mãos (*no contexto comum*), sei que se tenho mãos então não sou um BIV (*no contexto comum e no cético*); logo, sei que não sou um BIV (*no contexto cético*).

A outra vantagem da resposta contextualista é que ela explica a força do argumento cético onde as outras abordagens falham. A condicional *Se sei que ME, então sei que não-BIV* é verdadeira, pensa DeRose, em todos os contextos, pois pode ser descoberta por reflexão apenas. Não importa o quão alto seja o padrão num determinado contexto, sempre poderemos sabê-la. Além do mais, a nossa impossibilidade de responder ao cético se deve à mudança de contexto conversacional comum para um contexto no qual a dúvida cética é levantada; mas resguardados a contextos comuns, o ceticismo não nos ameaça.

O contextualismo enfrenta diversas objeções, das quais mencionarei apenas algumas. A mais óbvia é que o contextualista concede demais ao cético; basta que a dúvida cética seja introduzida num contexto para que não possamos atribuir conhecimento aos sujeitos. E mais; em contextos epistemológicos, em que é extremamente difícil ficar longe da dúvida cética, o contextualismo implica que todas as nossas atribuições de conhecimento são também falsas. E as coisas podem ainda piorar, pois se por descuido de meu pensamento longo acabo por formular a hipótese cética, deixo de ter conhecimento. O contextualismo parece ir longe demais. Para que S saiba que *p* é preciso que seu contexto esteja completamente livre da dúvida cética, o que certamente só ocorre com agentes pouquíssimos reflexivos. Para respostas breves a essas objeções e mais referências, veja HANNON 2016.

3. Coda

O ceticismo é sem dúvida um dos puzzles filosóficos mais fascinantes, não à toa tem ocupado a atenção dos filósofos pelo menos desde os gregos antigos. Respondê-lo, contudo, não tem sido tarefa fácil; o nosso breve exame já sinaliza para dificuldade de se atingir uma resposta satisfatória – e quem dirá definitiva! Há filósofos que, como Kant, acham um

“escândalo não haver uma prova do mundo externo”. Do meu lado, prefiro pensar que é um bem para filosofia que assim continue: a beleza dos problemas filosóficos reside certamente nos puzzles que os geram.

Bibliografia

- ADAMS, F. e CLARKE, M. (2005) “Resurrecting the tracking theories. *Australasian Journal of Philosophy*, 83 (2): 207-221.
- BAUMANN, P. (2008) “Was Moore a Moorian? On Moore and Scepticism”. *European Journal of Philosophy* 17 (2): 181-200.
- BECKER, K. (2007) *Epistemology Modalized*. Routledge.
- BECKER, K. e BLACK, T. (2012) *The Sensitivity Principle in Epistemology*. Cambridge University Press.
- BRIESER, J. (2010) “Reconsidering Closure, Underdetermination, and Infallibilism”. *Grazer Philosophische Studien* 80, 221-234.
- Brueckner, Anthony. 1986. Brains in a Vat. *Journal of Philosophy* 83: 148-67
- BRUECKNER, A. (1994) “The Structure of Skeptical Argument”. *Philosophy and Phenomenological Research* 54, 827-834.
- BRUECKNER, A. (2005) “Fallibilism, Underdetermination, and Skepticism”. *Philosophy and Phenomenological Research* 71, 384-391.
- BRUECKNER, A. (2010) “Ceticismo e Fechamento”. In *Investigação Filosófica*, v. 5, n.1, 2014. Tradução de L. H. Marques Segundo. URL: <http://periodicoinvestigacaoofilosofica.blogspot.com.br/2014/09/luiz-helvecio-marques-segundo-traducao.html>.
- COHEN, S. (1988) “How to be a Fallibilist”, *Philosophical Perspectives*, Volume 2: 91–123.
- COLLINS, J. M. (2016) “Epistemic Closure Principles”, *The Internet Encyclopedia of Philosophy*. URL = <http://www.iep.utm.edu/epis-clo/>.
- DESCARTES, R. (1641) *Meditações*. Coleção Os Pensadores.
- DeROSE, K. (1995) “Solving the Skeptical Problem”. *The Philosophical Review* 104 (1): 1-52. URL: <http://www.revistafundamento.ufop.br/index.php/fundamento/article/>

view/224/85.

DRETSKE, F. (1970) “Operadores Epistêmicos”. *Intuitio* v.2, n.3, 2009, 400-418. Tradução de Luis Fernando Munaretti da Rosa.

DRETSKE, F. (1971) “Conclusive Reasons”. *Australasian Journal of Philosophy* 49 (1): 1-22.

DRETSKE, F. (2005) “The Case against Closure,” in Steup and Sosa (eds.), 13-26.

GOLDMAN, A. (1976) “Discrimination and Perceptual Knowledge”. *Journal of Philosophy* 73: 771-791.

GOLDMAN, A. (1986) *Epistemology and Cognition*. Harvard University Press.

HANNON, M. J. (2016) “Skepticism and Contextualism”. In *Routledge Handbook of Epistemic Contextualism*, (Ed.) J. J. Ichikawa.

HELLER, M. (1999) “Closure and Relevant Alternatives”. *Australasian Journal of Philosophy*, 77(2): 196–208.

HICKEY, L. P. (2016) “The Brain in a Vat Argument”. *The Internet Encyclopedia of Philosophy*. URL = <http://www.iep.utm.edu/brainvat/>.

KRIPKE, S. (2011) “Nozick on Knowledge”. In *Philosophical Troubles: Collected Papers vol. 1*. Oxford University Press.

LUPER, S. (2016) “Epistemic Closure”, *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2016/entries/closure-epistemic/>.

MOORE, G. E. (1939) “A Proof of an external World”. [Trad. brasileira: “Prova de um mundo exterior”, em *Coleção Os Pensadores*, 1985].

NOZICK, R. (1981) *Philosophical Explanations*. Harvard University Press.

POLLOCK, J. e CRUZ, J. (1999) *Contemporary Theories of Knowledge*. Rowman & Littlefield.

PRITCHARD, D. (2002) “Recent Work on Radical Scepticism”. *American Philosophical Quarterly* 39, 215-57.

PRITCHARD, D. (2002) “Ressurrecting the Moorean Response to Scepticism”. *International Journal of Philosophical Studies* 10, 283-307.

PRITCHARD, D. (2005) “The Structure of Sceptical Arguments”. *Philosophical Quarterly* 55, 37-52.

PRITCHARD, D. (2007) "How to be a Neo-Moorean". In *Internalism and Externalism in Semantics and Epistemology*, (ed.) S. Golderberg. *Oxford University Press*.

PUTNAM, H. (1981) "Cérebros numa Cuba". In: *Fundamento*, n. 8, 2004. Tradução de L. H. Marques Segundo. URL: <http://www.revistafundamento.ufop.br/index.php/fundamento/article/view/217/76>.

SOSA, E. (1999) "How to defeat the opposite of Moore". *Philosophical Perspectives* 13, 137-49.

WARFIELD, D. (2004) "When Epistemic Closure Does and Does not Fail: a Lesson from the History of Epistemology," *Analysis* 64: 35-41.

O Enigma de Frege¹

Frege's Puzzle

Juliana Faccio Lima
University of California Santa Barbara

Resumo

Neste artigo irei apresentar o enigma de Frege e o problema que ele coloca para a Teoria da Referência Direta como uma teoria para o conteúdo semântico de nomes próprios. Discutirei como a Teoria Fregeana do Significado pode solucionar o enigma e algumas objeções levantadas a ela. Por fim, analisarei a teoria proposta por Nathan Salmon em Frege's Puzzle. Argumento que, apesar de ser uma resposta engenhosa do ponto de vista da Teoria da Referência Direta, ela tem problemas sérios e, portanto, devemos rejeitá-la como uma teoria apropriada para nomes próprios.

Palavras-chave

Enigma de Frege. Teoria da Referência Direta. Nathan Salmon.

1 - Gostaria de agradecer a Roberta Aguiar, Kevin Falvey, Iago Bozza Francisco e Sagid Salles pelas discussões e revisões que tornaram o presente artigo mais claro.

Abstract

In this paper I explain Frege's Puzzle and the problem it creates for the Direct Reference Theory as a theory for the semantic content proper names. I also discuss how the Fregean Theory of Reference may explain away the puzzle and some of the most common objections raised to it. At last, I consider Nathan Salmon's theory proposed in Frege's Puzzle. I argue that despite being a clever reply to the problem created by the puzzle on behalf of the Direct Reference Theory, it has serious flaws which should prompt us to reject it as an adequate theory for proper names.

Keywords

Frege's Puzzle. Direct Reference Theory. Nathan Salmon.

O nome 'Enigma de Frege' pode ser usado para se referir a dois fenômenos distintos mas relacionados que podem surgir quando consideramos o conteúdo de crenças e o valor da informação expressa por sentenças de identidade. Normalmente tomamos como verdadeira a sentença 'Lois Lane acredita que Superman voa'. Agora, quando substituímos 'Superman' por 'Clark Kent', a sentença resultante 'Lois Lane acredita que Clark Kent voa' nos parece falsa. Mas como é possível haver essa diferença no valor de verdade entre as duas sentenças se tanto 'Superman' e 'Clark Kent' se referem à mesma pessoa, e essa é a única diferença entre as duas sentenças? Além do mais, se Superman e Clark Kent são a mesma pessoa, como pode ser informativo dizer à Lois Lane 'Superman é Clark Kent.' mas não tão informativo dizer a ela que 'Superman é Superman.'?

As duas versões do enigma apresentadas acima foram originalmente apresentadas por Gottlob Frege em seu artigo 'Sobre o Sentido e a Referência' publicado em 1892. Nele Frege usa o enigma para argumentar a favor da sua teoria semântica (*Teoria Fregeana do Significado*), entre outras coisas.

Hoje em dia o Enigma de Frege é frequentemente usado como base para objeções à Teoria da Referência Direta, considerada a principal teoria rival à Teoria Fregeana do Significado. Neste artigo irei apresentar o enigma de Frege e o problema que ele coloca

para a Teoria da Referência Direta como um teoria para o conteúdo semântico de nomes próprios. Discutirei como a Teoria Fregeana do Significado pode solucionar o enigma e algumas objeções levantadas a ela. Por fim, analisarei a teoria proposta por Nathan Salmon em Frege's Puzzle. Argumento que, apesar de ser uma resposta engenhosa do ponto de vista da Teoria da Referência Direta, ela tem problemas sérios e, portanto, devemos rejeitá-la como uma teoria apropriada para nomes próprios.

1. Preliminares

Antes de apresentar o enigma, vou explicar alguns termos técnicos e o propósito geral da Teoria Fregeana do Significado e da Teoria da Referência Direta com relação a nomes próprios.

Nomes próprios são, grosso modo, aquelas expressões linguísticas comumente chamados de 'nomes', como 'Aline', 'Juliana', 'Bill Gates', 'Superman', 'Clark Kent', etc. As duas teorias a ser discutidas aqui classificam nomes próprios como *termos singulares*. Termos singulares são expressões linguísticas cuja função é se referir a um, e apenas um, objeto. Por exemplo, o nome 'Juliana' tem a função de se referir a mim, o nome 'Bill Gates' tem a função de se referir a Bill Gates, etc.

Além de nomes próprios, expressões *indexicais* também são usualmente consideradas termos singulares. Indexicais são expressões linguísticas cujo referente pode variar sistematicamente de acordo com o contexto em que é usado. A expressão 'eu' é uma expressão indexical pois o objeto ao qual 'eu' se refere depende de quem profere a expressão, não é "fixo" como no caso de nomes próprios. Por exemplo, suponha que estamos todos em um bar bebendo cerveja quando o garçom chega na nossa mesa e pergunta quem quer mais cerveja. Quando Juliana responde dizendo 'Eu quero cerveja.', o uso de 'eu' se refere à Juliana; e ela está dizendo que Juliana quer cerveja. Quando Aline responde dizendo 'Eu quero cerveja.', ela obviamente não diz que Juliana quer cerveja, mas que Aline quer cerveja. Nesse caso, 'eu' se refere a Aline. Outras expressões indexicais são 'isso', 'ele', 'hoje', 'agora', 'aqui', etc.

Existe ainda um terceiro tipo de expressões linguísticas, as *descrições definidas*, que também pode ser classificado como termo singular dependendo da teoria. Descrição definidas são expressões com a seguinte forma 'o *F*' ou 'a *F*', tal que '*F*' representa uma expressão predicativa. Exemplos de descrições definidas são: 'o presidente dos Estados Unidos em 2012', 'o inventor do zíper', 'a capital do Brasil', etc.

Aqui eu vou focar nas versões do Enigma de Frege apenas com relação a nomes próprios. Mas é importante notar que há versões do Enigma para indexicais e descrições definidas. Fica como tarefa para o leitor ao final do artigo pensar como o mesmo fenômeno pode aparecer com esses dois tipos de expressões linguísticas.

A Teoria Fregeana do Significado e a Teoria da Referência Direta discordam a respeito da natureza do *conteúdo semântico* de nomes próprios. O *conteúdo semântico* de um nome próprio é o conteúdo expresso por ele que contribui para o conteúdo semântico expresso por sentenças declarativas. Sentenças declarativas são aquelas sentenças que podem ser verdadeiras ou falsas, como 'Juliana é brasileira.' - por oposição a sentenças que não são verdadeiras ou falsas, por exemplo 'Qual a nacionalidade de Juliana?'. Por poderem ser verdadeiras ou falsas, sentenças declarativas ('sentenças' daqui em diante) têm *condições de verdade* associadas a elas. Quando essas condições são satisfeitas, dizemos que a sentença é verdadeira; quando não satisfeitas, dizemos que a sentença é falsa. Para os propósitos desse artigo, as condições de verdade associadas a uma sentença vão ser identificadas com o *conteúdo semântico* ou a *proposição* expressa pela sentença.

Assim como podemos entender que sentenças são determinadas pelas expressões básicas que as compõem, também podemos entender que o conteúdo semântico expresso por elas é determinado pelo conteúdo semântico expresso pelas suas expressões básicas. Por uma série de motivos que não vou discutir aqui, é natural entender que a sentença 'Juliana é brasileira.' é constituída por duas expressões básicas: o nome, 'Juliana', e o predicado 'ser brasileira'. Da mesma forma, algumas teorias semânticas propõem que o conteúdo semântico (proposição ou condições de verdade) expresso por 'Juliana é brasileira.' é determinado pelo conteúdo semântico de 'Juliana' e 'ser brasileira'. O princípio por trás dessa ideia é conhecido como *Princípio de Composicionalidade* com relação à proposição ou conteúdo semântico expresso por uma sentença declarativa:

(Tese 1) O conteúdo semântico (ou proposição) expresso por uma sentença declarativa é determinado pelo conteúdo semântico das expressões básicas que compõem a sentença e a ordem em que aparecem na sentença.

Com isso, podemos dizer um pouco mais a respeito do conteúdo semântico de nomes próprios: ele é a contribuição que o nome dá às condições de verdade da sentença da qual ele faz parte.

O leitor atento já deve ter percebido que também precisamos de uma teoria sobre o conteúdo semântico de predicados. As teorias das quais falarei aqui também falam sobre o conteúdo semântico de predicados. E também há uma versão do Enigma de Frege envolvendo predicados. Mas vou deixar essa parte das teorias de lado porque, como disse anteriormente, vou focar a discussão no conteúdo semântico de nomes próprios apenas.

Antes de seguir adiante, quero deixar claro quais dos conteúdos associados a uma sentença é seu conteúdo semântico. Quando uma pessoa profere uma sentença, ela possivelmente expressa uma série de conteúdos. Mas nem todos são o conteúdo semântico da sentença usada. Por exemplo, quando digo em voz alta ‘Aline é brasileira.’, ao mesmo tempo que expresso o conteúdo semântico dessa sentença, ou suas condições de verdade, também posso expressar certas emoções dependendo do meu tom de voz, minhas expressões faciais e corporais. Se a mãe da Aline me pergunta qual é a nacionalidade da Aline, eu posso responder com uma certa expressão facial (sobrancelhas mais juntas, olhos um pouco mais fechados que o normal, o canto esquerdo da boca meio puxado para o lado) ‘Aline é brasileira.’ Nesse caso, muito possivelmente vou passar não só a informação que Aline é brasileira como também a informação que suspeito que tem alguma coisa estranha com a mãe de Aline, afinal uma situação onde uma mãe não sabe a nacionalidade de sua própria filha é bastante inusitada. Teorias sobre o conteúdo de expressões linguísticas estão interessadas apenas no primeiro tipo de informação, pois este é o conteúdo semântico. O segundo tipo de informação é classificado como *conteúdo pragmático*.

Um modo intuitivo que temos para saber se um conteúdo que associamos a uma sentença é semântico ou pragmático, é nos perguntar se esse conteúdo é sistematicamente expreso pela sentença ou não. O conteúdo que eu estou suspeitando de algo errado com a mãe da Aline não é um conteúdo sistematicamente associado com a sentença ‘Aline é brasileira.’ Afinal, ao proferir a mesma sentença em outros contextos, digamos quando Valéria me pergunta qual a nacionalidade de Aline, eu não passo a informação de que há algo de errado com a mãe da Aline, mas apenas a informação que Aline é brasileira. A primeira informação é um conteúdo comunicado apenas no contexto envolvendo a mãe da Aline. Isso quer dizer, entre outras coisas, que minhas suspeitas de que há algo de errado com a mãe da Aline *não* é relevante para as condições de verdade daquilo que expressei quando disse ‘Aline é brasileira.’ Portanto essa informação não é faz parte do conteúdo semântico expreso por um proferimento de ‘Aline é brasileira.’

Claro, esse critério não é aceito por todos como sendo um critério bom para fazer a distinção em todos (ou até em alguns) casos. Mas ele funciona relativamente bem para uma primeira abordagem da questão, que é meu propósito aqui.

2. Teoriada Referência Direta

Como normalmente as duas versões do Enigma de Frege estão ligadas a objeção à Teoria da Referência Direta, primeiro vou apresentar a teoria e depois os Enigmas de Frege.

O nome ‘Teoria da Referência Direta’ vai ser usado aqui como um termo guarda-chuva. Existem inúmeras teorias que podem ser classificadas como Teoria da Referência Direta; basicamente, uma teoria para cada filósofo que se diz defensor da Teoria da Referência Direta. O que eu aqui chamo de ‘Teoria da Referência Direta’ se refere a qualquer teoria que se compromete com a seguinte tese (que vou também chamar de ‘tese característica da Teoria da Referência Direta’), dentre outras:

(Tese 2) O conteúdo semântico de nomes próprios é unicamente o objeto referido pelo nome próprio.



No caso de ‘Aline’, o objeto referido pelo nome é a própria Aline. Logo, a própria Aline é a única contribuição para a proposição expressa pela sentença ‘Aline é brasileira.’

Como expliquei anteriormente, predicados também têm conteúdo semântico. Mas não vou entrar na discussão a respeito de qual é seu conteúdo semântico. Apenas por uma questão de consistência, vou supor que para a Teoria da Referência Direta o conteúdo de predicados é também o seu referente. E vou supor que o referente de um predicado é uma propriedade, o que quer que ela seja. Portanto, no caso do predicado ‘ser brasileira’, sua contribuição semântica é a propriedade de ser brasileira.

Do que foi dito nos últimos dois parágrafos temos que a proposição expressa por ‘Aline é brasileira.’ é constituída por Aline e a propriedade de ser brasileira. Usualmente proposições são representados por n-tuplas ordenadas contendo o conteúdo semântico das expressões básicas que compõem a sentença. Seguindo essa convenção, a proposição expressa por ‘Aline é brasileira.’ é $\langle Aline, ser\ brasileira \rangle$, tal que ‘Aline’ representa Aline ela mesma, e ‘*ser brasileira*’, a propriedade de ser brasileira. O que isso quer dizer é que para a Teoria da Referência Direta, ‘Aline é brasileira.’ é verdadeira se, e somente se, Aline tem a propriedade de ser brasileira.

Uma consequência da conjunção da Tese 1 (Princípio de Composicionalidade) e da Tese 2 (tese característica da teoria da referência direta) é:

(Tese 3) Duas sentenças com diferentes nomes co-referenciais (ou seja, nomes diferentes do mesmo objeto) mas que são de outra forma idênticas expressam a mesma proposição.

Um par de sentenças que exemplifica essa tese é: ‘Superman voa.’ e ‘Clark Kent voa.’ Essas duas sentenças são idênticas exceto pelo fato que uma contém o nome ‘Superman’ e a segunda, o nome ‘Clark Kent’, que são diferentes nomes do mesmo objeto. Eles são diferentes nomes no sentido de serem diferentes símbolos, diferentes sequências de letras; o primeiro começa com ‘S’ enquanto que o segundo começa com ‘C’. Mas eles são co-referenciais porque se referem ao mesmo objeto, a saber, , tal que  representa Superman (ou Clark Kent) ele mesmo.

Com isso, é fácil ver porque as sentenças acima expressam a mesma proposição de acordo com a Tese 1 e a Tese 2. A proposição expressa por uma sentença é constituída pelos referentes das expressões que a compõe. Como os referentes das expressões que compõem ‘Superman voa.’ e ‘Clark Kent voa.’ são os mesmos, a proposição expressa por elas também é a mesma, a saber, $\langle \text{stick figure with red triangle}, \text{voar} \rangle$, tal que ‘voar’ representa a propriedade de voar. Uma das coisas que isso quer dizer é que ‘Superman voa.’ e ‘Clark Kent voa.’ sempre terão as mesmas condições de verdade e, consequentemente, o mesmo valor de verdade.

3. Primeira versão do Enigma de Frege

Um dos dois fenômenos chamados de ‘Enigma de Frege’ se torna saliente quando analisamos com cuidado uma das implicações da Tese 3 que mencionei no final da seção anterior - uma versão restrita a nomes próprios do princípio mais geral chamado ‘Princípio de Substituição *salva veritate*’:

(Tese 4) A substituição de um nome em uma sentença por um nome diferente mas co-referencial não altera o valor de verdade da proposição expressa pela sentença.

De acordo com a Tese 4, o nome ‘Superman’ em ‘Superman voa.’ pode ser substituído por ‘Clark Kent’, cujo resultado é ‘Clark Kent voa.’, sem que o valor de verdade da última sentença seja diferente do valor de verdade da sentença original. Essa tese é verdadeira na Teoria da Referência Direta justamente porque ela defende que a única contribuição semântica de nomes próprios é o seu próprio referente. Então não importa se usamos ‘Superman’ ou ‘Clark Kent’, expressaremos a mesma proposição, ou seja, as mesmas condições de verdade. Portanto, se ‘Superman voa.’ expressa uma proposição verdadeira, então ‘Clark Kent voa.’ - o resultado da substituição de ‘Superman’ por ‘Clark Kent’ - também é verdadeira.

No caso das sentenças que usei para ilustrar o princípio, a Teoria da Referência Direta não aparenta ter problemas. Em certa medida é intuitivo que ‘Superman voa.’ tenha sempre o mesmo valor de verdade que ‘Clark Kent voa.’. Entretanto, existem pares de sentenças que aparentemente violam e, conseqüentemente, falsificam a tese 4. Considere a sentença ‘Lois Lane acredita que Superman voa.’ De acordo com a estória tal como escrita por Jerry Siegel, dizemos que sentença é verdadeira porque a proposição expressa por ela é verdadeira. Pela Tese 4 e do fato que ‘Superman’ e ‘Clark Kent’ são nomes co-referenciais, deveríamos poder substituir o primeiro pelo segundo, e a sentença resultante seria verdadeira -o mesmo valor de verdade que a sentença original. Contudo, intuitivamente, não é isso que acontece. Fazendo a substituição dos nomes ficamos com a sentença ‘Lois Lane acredita que Clark Kent voa.’ E de acordo com a estória, não é o caso que Lois Lane acredita que Clark Kent voa. Onde se segue que esta última sentença é, em realidade, falsa, contrário ao que é afirmado na Tese 4.

Aqui, então, temos um exemplo da primeira versão do Enigma de Frege: como é possível que as sentenças ‘Lois Lane acredita que Superman voa.’ e ‘Lois Lane acredita que Clark Kent voa.’ tenham valores de verdade diferentes sendo que ‘Superman’ e ‘Clark Kent’ são co-referenciais? Aparentemente esse fenômeno suporta a ideia que existe alguma outra coisa além do referente de nomes próprios, que é relevante para o valor de verdade de sentenças. Mas o que mais poderia ser relevante para o valor de verdade de sentenças?

Essa versão do enigma de Frege, ou o fenômeno a ser explicado, não aparece apenas no caso do verbo *acreditar*. Na verdade ele pode aparecer em quaisquer pares de sentença contendo verbos que expressam *atitudes proposicionais* com nomes co-referenciais em seu escopo. Verbos que expressam atitudes proposicionais são aqueles que expressam uma relação entre um sujeito e uma proposição. No caso de ‘Lois Lane acredita que Superman voa.’, ‘acreditar’ é o verbo que expressa uma atitude proposicional, e ele relaciona Lois Lane (referido por ‘Lois Lane’) com a proposição $\langle \text{ Superman voa. } \rangle$ (expressa por ‘que Superman voa’).

Aqui, *acreditar* é a atitude que Lois Lane têm com relação à proposição $\langle \text{voar} \rangle$. Outros verbos que expressam atitudes proposicionais são: ‘desejar’, ‘saber’, e (em geral) outros verbos seguidos de ‘que’ e uma sentença declarativa. Todos eles podem, a princípio, gerar um exemplo do fenômeno explicado acima: ‘Lois Lane *deseja* que Superman conceda uma entrevista.’ é verdadeira mas ‘Lois Lane *deseja* que Clark Kent conceda um entrevista.’ é falsa; ‘Lois Lane *sabe* que Clark Kent trabalha no Daily Planet’ é verdadeira mas ‘Lois Lane *sabe* que Superman trabalha no Daily Planet’ é falsa; etc.

De forma geral, o fenômeno a ser explicado pela primeira versão do enigma de Frege é: suponha que ‘cog’ represente um verbo de atitude proposicional (*acreditar*, *desejar*, etc.) e ‘P’ e ‘P’” sentenças declarativas cuja única diferença é possuírem distintos nomes co-referenciais. Como é possível que pares de sentenças do tipo ‘S cog P’ e ‘S cog P’” tenham diferentes valores de verdade?

Essa versão do Enigma de Frege gera um problema sério para a Teoria da Referência Direta. Relembremos que a Tese 4 é implicada pela Tese 1 (Princípio de Composicionalidade) junto com a Tese 2 (tese característica da teoria). Entretanto o Enigma sugere que a Tese 4 é falsa. Logo, para evitar o comprometimento com essa falsa tese, precisamos rejeitar a Tese 1 ou a Tese 2. Mas a Tese 1 é aparentemente não problemática. Portanto, se os contra exemplos à Tese 4 realmente existem, só nos resta rejeitar a Tese 2, a tese característica da teoria da referência direta. Em outras palavras, se tais exemplos existem, então nos resta apenas rejeitar a Teoria da Referência Direta.


4. Segunda versão do Enigma de Frege

A segunda versão do Enigma de Frege está de alguma forma relacionada com a Tese 4 mas diz respeito não ao valor de verdade de sentenças mas sim ao seu *valor cognitivo* ou *informatividade*. Essa versão do enigma é a que encontramos logo no começo do artigo de Frege.

Frege argumenta que sentenças cuja forma gramatical é ‘a é a’ e sentenças como ‘a é b’ têm diferentes valores informativos, ainda que ‘a’ e ‘b’ sejam nomes próprios que se refiram ao mesmo objeto². Por exemplo, se Lois Lane passa a acreditar no que é expresso

2 - Em seu artigo, Frege oferece exemplos de pares de sentença onde o enigma surge envolvendo descrições definidas. Aqui eu vou alterar o exemplo de Frege e usar apenas nomes próprios porque atualmente é mais comum encontrar teorias que aceitam que o enigma surge no caso de nomes próprios do que teorias que aceitem que o enigma surge no caso de descrições definidas. A razão disso é que grande parte das teorias semânticas não classificam descrições definidas como termos singulares como Frege defendia. E o Enigma de Frege só surge no caso de descrições definidas se as considerarmos termos singulares.

por ‘Superman é Superman,’ ela não adquire nenhum conhecimento relevante sobre o mundo. Além disso, essa sentença é apenas uma instância do princípio de identidade - todo objeto é idêntico a si mesmo -, e sabemos da sua verdade de forma *a priori*. Por outro lado, o mesmo não é verdade da sentença ‘Superman é Clark Kent.’ Esta última é informativa em um sentido em que a primeira não é. Se Lois Lane começar a acreditar no conteúdo expresso por essa sentença, haverá uma alteração significativa no seu sistema de crenças: todas aquelas propriedades e relações que ela acreditava pertencer a Superman passarão a ser atribuídas também a Clark Kent, e vice-versa. Além do mais, é razoável supor que o conhecimento de que Superman é Clark Kent depende de investigações empíricas, e é difícil ver como poderia ser obtido de outra forma, quer dizer, sem investigações empíricas. Portanto, sabemos da identidade ‘Superman é Clark Kent.’ de forma *a posteriori*.

O enigma nesse caso é: como é possível que sentenças do tipo ‘a é a’ sejam conhecíveis *a priori* e não informativas³, enquanto que sentenças como ‘a é b’ são conhecíveis *a posteriori* e informativas, sendo que ‘a’ e ‘b’ são termos singulares *co-referenciais*? Dado que ‘Superman’ e ‘Clark Kent’ se referem à mesma pessoa, a saber, , por que ‘Superman é Superman.’ não é informativa e é conhecida sem qualquer investigação enquanto que ‘Superman é Clark Kent.’ é informativa e demanda investigação? Como explicar a diferença com relação à informatividade e modo de conhecimento do conteúdo dessas duas sentenças se o referente dos nomes é o mesmo?

Essa versão do enigma de Frege também pode ser usada como base para uma objeção à Teoria da Referência Direta. De acordo com essa teoria, a proposição expressa pelas sentenças ‘Superman é Superman.’ e ‘Superman é Clark Kent.’ é a mesma, a saber, $\langle \text{stick figure}, \text{stick figure} \rangle$, *identidade*, $\langle \text{stick figure}, \text{stick figure} \rangle$, tal que ‘*identidade*’ representa a relação de identidade entre dois objetos. Ademais, se o que se supõe ser duas proposições são na verdade uma e a mesma, então elas deveriam ter as mesmas propriedades epistêmicas (ser conhecível *a priori* ou *a posteriori*) e valores informativos (ser informativa ou não informativa). Contudo, isso não é o caso na Teoria da Referência Direta porque, intuitivamente, a primeira é *a priori* e não informativa, mas a segunda é *a posteriori* e informativa. Portanto, se o fenômeno descrito aqui é legítimo, então devemos rejeitar a Teoria da Referência Direta, afinal ela não explica como é possível as duas sentenças terem propriedades diferentes.

Existem algumas alternativas disponíveis para salvar a teoria da referência direta do problema posto pelas duas versões do Enigma de Frege - claro, nem todas são igualmente

3 - Quando digo que sentenças do tipo ‘a é a’ são não informativas quero dizer que elas não são informativas no mesmo sentido em que sentenças do tipo “a é b” são. Mas com isso não quero comprometer Frege com a afirmação que sentenças do tipo ‘a é a’ não tem nenhum valor informativo. O que Frege de fato defende é que sentenças desses dois tipos têm valores informativos diferentes. Mas é possível que Frege pensasse que tais sentenças tenham algum valor informativo, ainda que diferente do valor informativo de sentenças como “a é b”.

satisfatórias. Elas vão desde pequenas alterações no modo de entender algumas teses – como as teorias propostas por Salmon – até alterações radicais que envolvem a rejeição da Tese 2, a tese característica da teoria da referência direta – como a teoria proposta por Frege. Nas seções seguintes irei apresentar duas soluções para as duas versões do Enigma de Frege.

5. Teoria Fregeana do Significado

Já que o enigma foi proposto por Frege, primeiro apresentarei sua explicação do fenômeno e a resposta ao problema gerado para a teoria da referência direta. De modo geral, a ideia de Frege é a seguinte: já que os pares de sentença ‘Lois Lane acredita que Superman voa.’/‘Lois Lane acredita que Clark Kent voa.’ e ‘Superman é Superman.’/‘Superman é Clark Kent.’ têm propriedades diferentes, então elas expressam proposições diferentes - e não a mesma proposição como afirma a Teoria da Referência Direta. Com isso, Frege rejeita a Tese 2, a tese característica da Teoria da Referência Direta que afirma que o conteúdo semântico de um nome próprio é apenas o referente. Ao invés disso, ele sustenta que:

(Tese 2*) O conteúdo semântico de nomes próprios é constituído por um *sentido* e um referente (se houver).

Quer dizer, Frege sustenta que, além do referente, o conteúdo semântico de um nome também inclui um sentido que é um modo de apresentação do seu referente.

O *sentido* de uma expressão linguística é, para Frege, aquilo que um falante competente naquele idioma apreende quando ele entende o significado da expressão linguística. Por exemplo, quando um falante competente na língua portuguesa entende o nome ‘Superman’, ele apreende um certo conteúdo, que não é Superman ele mesmo, mas um conteúdo cognitivo que *apresenta* e é satisfeito por Superman. Esse conteúdo é o que Frege chama de ‘sentido’.

Em uma famosa nota de rodapé, Frege oferece um exemplo de como ele entende o sentido de um nome (p. 159, nota 4, tradução minha):

No caso de um nome próprio real como ‘Aristóteles’, opiniões a respeito do [seu] sentido podem diferir. Ele pode ser, por exemplo, tomado como o seguinte: o pupilo de Platão e professor de Alexandre o Grande. Qualquer pessoa que associe esse sentido irá associar um sentido diferente à sentença ‘Aristóteles nasceu em Estagira.’ ao sentido que será associado por alguém que toma [o seguinte sentido] como sentido do nome [‘Aristóteles’]: o professor de Alexandre o Grande que nasceu em Estagira. Contanto que o referente se mantenha o mesmo, tais variações no sentido podem ser toleradas, apesar de que elas devem ser evitadas na estrutura teórica de ciências demonstrativas [como matemática e outras ciências] e não devem ocorrer numa linguagem perfeita.

Um modo comum de entender uma das coisas que Frege diz nessa nota de rodapé é que o sentido de um nome próprio deva ser identificado com o sentido de uma descrição definida⁴. Nesse caso, um possível sentido expresso por ‘Superman’ é *a pessoa que veste cuecas vermelhas por cima da calça*; e para ‘Clark Kent’, o sentido pode ser *o jornalista do Planeta Diário apaixonado por Lois Lane*. Já no caso de ‘Aline’, o sentido pode ser *a aluna de doutorado da UFPR que estuda Wittgenstein*. E assim por diante. Exatamente qual sentido uma pessoa irá apreender quando entender o nome ‘Superman’, ‘Clark Kent’, ‘Aline’, etc. pode variar dependendo das outras crenças que ela tem do referente do nome. Os sentidos apresentados aqui são apenas sugestões.

É importante ressaltar que apesar de Frege admitir uma variação no sentido apreendido por diferentes sujeitos quando eles entendem um nome, ele acredita que o sentido é *objetivo*. Isso quer dizer que um sentido deve ser, pelo menos a princípio, capaz de ser apreendido por qualquer pessoa. Por exemplo, suponha que quando Aline entende o nome ‘Superman’, ela apreende o sentido *a pessoa que veste cuecas vermelhas por cima da calça*. Como sentidos são objetivos, então isso quer dizer que a princípio é possível que eu apreenda o mesmo sentido que Aline ainda que eu não apreenda o mesmo sentido nesse momento – suponha que nesse momento eu apreendo o sentido *o ser ficcional que voa criado por John Siegel*. Portanto, o sentido não pode ser identificado com sentimentos ou imagens mentais que Aline tem de Superman, pois esses não são compartilháveis. Isso quer dizer, entre outras coisas, que o sentido de ‘Superman’ que Aline apreende não pode ser identificado com a “foto” mental de Superman que ela tem na memória. “Fotos” mentais ou esse “filme” que passa na nossa mente quando temos experiências do mundo através dos nossos sentidos

4 - Para alguns filósofos, não é claro que Frege de fato quis dizer que o sentido de um nome é o sentido de uma descrição definida. Para interpretações alternativas, ver Dummett (1981, 1991) e Evans (1982)

não são acessíveis a ninguém mais além da pessoa que as têm.

Tudo o que foi dito a respeito de nomes próprios pode ser estendido para predicados. Eles têm um sentido que é um modo de apresentação do seu referente, uma propriedade (se houver). Mas essa discussão vai ser deixada de lado.

Agora que sabemos em linhas gerais o que é um sentido, podemos compreender melhor a Teoria Fregeana do Significado. Frege aceita a Tese 1: o conteúdo semântico expresso por uma sentença é determinado pelo conteúdo semântico expresso por suas expressões linguísticas básicas. Mas diferente da Teoria da Referência Direta, Frege acredita que esse conteúdo expresso por uma sentença é determinado pelos sentidos expressos por suas expressões básicas e não só seus referentes. No caso do conteúdo semântico expresso pela sentença ‘Superman voa,’ o sentido é determinado pelo sentido ‘Superman’ e pelo sentido de ‘voar’: $\langle mda_{SM}, mda_{voa} \rangle$, tal que ‘ mda_{SM} ’ representa o sentido *a pessoa que veste cuecas vermelhas por cima da calça*, e ‘ mda_{voa} ’ representa o sentido de ‘voar’. Já o seu referente é determinado pelos referentes de ‘Superman’ e ‘voar’. Para Frege, e aqui temos uma afirmação bastante contra-intuitiva, o referente de sentenças são *valores de verdade*. Frege admite apenas dois diferentes valores de verdade: o Verdadeiro e o Falso. Portanto, todas as sentenças verdadeiras têm o mesmo referente, a saber, o Verdadeiro. Tanto ‘Superman voa,’ ‘Superman é Superman.’ quanto ‘ $2 + 2 = 4$ ’ e ‘A neve é branca.’ se referem ao mesmo objeto, apesar de terem sentidos diferentes. E todas as sentenças falsas têm o mesmo referente, a saber, o Falso.

Antes de continuar a explicação de como a teoria de Frege soluciona a segunda versão do Enigma de Frege, eu gostaria de fazer uma clarificação terminológica. Apesar de Frege defender que o conteúdo semântico de uma sentença é constituído pelo seu sentido e pelo seu referente, o que hoje em dia chamamos de ‘proposição’ usualmente se refere apenas ao que é o *sentido* da sentença na teoria de Frege. Quando eu usar as expressões ‘conteúdo semântico’ ou ‘proposição’ daqui em diante para falar da teoria de Frege, estarei me referindo apenas ao sentido de uma sentença. Quando o referente da sentença for relevante para discussão, irei me referir a ele explicitamente como ‘referente da sentença’ e não como ‘conteúdo semântico da sentença’, apesar de Frege considerar o referente da sentença uma parte do seu conteúdo semântico.

Com o que foi explicado da teoria de Frege até agora já podemos ver sua solução para a segunda versão do Enigma de Frege. O enigma nesse caso era como explicar que as proposições expressas por ‘Superman é Superman.’ e ‘Superman é Clark Kent.’ têm diferentes propriedades epistêmicas (a primeira é *conhecível a priori* enquanto a segunda não) e valor informativo (a primeira não é informativa enquanto a segunda é). De acordo

com a Teoria Fregeana, não há nenhum mistério nessa diferença.

As duas sentenças de identidade expressam proposições diferentes para Frege porque pelo menos uma de suas expressões básicas têm sentidos diferentes, a saber, 'Superman' e 'Clark Kent'.

Como vimos anteriormente, 'Superman é Superman.' expressa a proposição $\langle mda_{SM} \text{ identidade}, mda_{SM} \rangle$ - «a pessoa que veste cuecas vermelhas por cima da calça é a pessoa que veste cuecas vermelhas por cima da calça». Mas a sentença 'Superman é Clark Kent', expressa uma proposição diferente: $\langle mda_{SM} \text{ identidade}, mda_{CK} \rangle$, tal que ' mda_{CK} ' representa o sentido o jornalista do Planeta Diário apaixonado por Lois Lane, que é diferente de mda_{SM} - «a pessoa que veste cuecas vermelhas por cima da calça é o jornalista do Planeta Diário apaixonado por Lois Lane». Mas como 'Superman' e 'Clark Kent' possuem o mesmo referente, o referente das sentenças 'Superman é Superman.' e 'Superman é Clark Kent.' se mantém o mesmo, ou seja, ambas terão necessariamente o mesmo valor de verdade.


Em suma, a razão pela qual as sentenças de identidade aqui analisadas diferem com relação a propriedades epistêmicas e informativas é que seus conteúdos semânticos são diferentes. Assim, o primeiro Enigma desaparece.

Com o que temos até agora ainda não é possível explicar a diferença no valor de verdade das sentenças 'Lois Lane acredita que Superman voa.' e 'Lois Lane acredita que Clark Kent voa.' na teoria de Frege. Como o valor de verdade de uma sentença é determinado pelo referente das expressões que a compõe, para permitir uma diferença no valor de verdade dessas sentenças, precisamos garantir que os referentes de 'Superman' e 'Clark Kent' sejam diferentes. Mas o que temos da teoria de Frege até agora é apenas a garantia que seus sentidos são diferentes, e diferentes sentidos não necessariamente garantem diferentes referentes.

Para assegurar que os referentes de 'Superman' e 'Clark Kent' nas sentenças acima sejam diferentes, Frege propõe a seguinte tese:

(Tese 5) Verbos que expressam atitudes proposicionais criam contextos indiretos.



Contextos indiretos são contextos criados por certas expressões, como aquelas que expressam atitudes proposicionais, que mudam o conteúdo semântico das expressões que aparecem no seu escopo. Por exemplo, em contextos *transparentes* (quer dizer, não indiretos), como em 'Superman voa.', o nome 'Superman' expressa um sentido mda_{SM} (também

chamado de ‘sentido ordinário’) e se refere a . Em contextos indiretos, por outro lado, o sentido expresso por ‘Superman’ muda. Quando no escopo de apenas uma expressão que gera contextos indiretos, o sentido expresso por ‘Superman’ muda para mda_{SM}^1 e o referente passa a ser mda_{SM} o sentido em contextos transparentes. Esse é o caso quando o nome ‘Superman’ é usado em ‘Lois Lane acredita que Superman voa.’ Se adicionarmos outro verbo que expressa atitudes proposicionais a esta última sentença, como em ‘Aline sabe que Lois Lane acredita que Superman voa.’, então ‘Superman’ passa a expressar um outro sentido mda_{SM}^2 , que é diferente tanto de mda_{SM}^1 quanto mda_{SM} , e se refere ao sentido mda_{SM}^1 . E se adicionarmos ainda outro verbo que expressa atitude proposicional, então ‘Superman’ passa a expressar o sentido mda_{SM}^3 , que é diferente de todos os outros mencionados até agora, e se refere a mda_{SM}^2 . E assim sucessivamente.

Com isso podemos finalmente solucionar a primeira versão do Enigma. Lembremos que o enigma é o de explicar como é possível que as sentenças ‘Lois Lane acredita que Superman voa.’ e ‘Lois Lane acredita que Clark Kent voa.’ possuem valores de verdade diferentes sendo que ‘Superman’ e ‘Clark Kent’ são nomes co-referenciais. Na Teoria Fregeana, esses nomes não são mais co-referenciais quando em contextos indiretos. Em ‘Lois Lane acredita que Superman voa.’, o nome ‘Superman’ expressa o sentido mda_{SM}^1 e se refere a mda_{SM} . Em ‘Lois Lane acredita que Clark Kent voa.’, o nome ‘Clark Kent’ expressa o sentido mda_{CK}^1 – modo de apresentação do sentido ordinário do nome - e se refere a mda_{CK} .

Nesse caso, a proposição expressa pela primeira sentença é $\langle mda_{LL}, mda_{acreditar}, \langle mda_{SM}^1, mda_{voa}^1 \rangle \rangle$, sendo que mda_{LL} , $mda_{acreditar}$, mda_{voa}^1 são sentidos que expressam modos de apresentação de Lois Lane, da relação de acreditar e do sentido ordinário de ‘voar’, respectivamente. E ela é verdadeira se os objetos que satisfazem os sentidos estão relacionados da maneira representada na proposição. Ou seja, se Lois Lane acredita em $\langle mda_{SM}^1, mda_{voa}^1 \rangle$, a proposição expressa pelo sentido ordinário de ‘Superman voa.’). Como isso é o caso de acordo com a estória, então ‘Lois Lane acredita que Superman voa.’ é verdadeira.

A proposição expressa pela segunda sentença é $\langle mda_{LL}, mda_{acreditar}, \langle mda_{CK}^1, mda_{voa}^1 \rangle \rangle$. E ela é verdadeira se Lois Lane acredita em $\langle mda_{CK}^1, mda_{voa}^1 \rangle$, a proposição expressa pelo sentido ordinário de ‘Clark Kent voa.’. Como isso não é o caso de acordo com a estória, então ‘Lois Lane acredita que Clark Kent voa.’ é falsa.

Na Teoria Fregeana do Significado, as duas sentenças que reportam crenças da Lois Lane podem, então, ter dois valores de verdade diferentes porque expressam condições de verdade diferentes satisfeitas por objetos diferentes. Quando Lois Lane pensa em  como *a pessoa que veste cuecas vermelhas por cima da calça*, ela acredita que ele voa. Mas quando ela pensa em  como *o jornalista do Planeta Diário apaixonado por Lois Lane*, ela não acredita



que ele voa. A razão disso? Lois Lane não sabe que Superman é Clark Kent. Essa é uma explicação não apenas aparentemente consistente, como também bastante intuitiva para a diferença no valor de verdade das sentenças.

6. Problemas para a solução de Frege





A solução de Frege para as duas versões do enigma parece ser bem atrativa. Mas não devemos aceitá-la sem antes examinarmos alguns dos problemas levantados na literatura. Aqui irei apresentar três problemas intimamente ligados com a solução para os Enigma: a dificuldade de Frege de explicar o que é um sentido; a geração de uma cadeia infinita de sentidos e o problema gerado pela variação de sentido de um nome. Esses não são necessariamente os principais problemas da teoria fregeana como um todo. A teoria tem outros problemas não diretamente relacionados com os enigmas discutidos aqui, como os conhecidos Argumento Modal, Argumento Semântico e o Argumento Epistemológico – famosamente apresentados por Saul Kripke (1980) e extensamente discutidos.

6.1. O que é um sentido?

A solução de Frege depende substancialmente de *sentidos* de expressões linguísticas que são *modos de apresentação* de objetos referidos por elas. Para avaliarmos sua solução, precisamos primeiro saber o que ele entende por sentido para então verificar se é realmente possível que haja sentidos e que eles tenham o papel que lhes é atribuído. Portanto, é razoável exigirmos de Frege uma explicação detalhada dessas entidades. Aqui, então, temos o primeiro grande problema para sua teoria.

Em seu artigo, Frege não explica em detalhes o que são sentidos. Como vimos anteriormente, ele afirma que sentidos são a) objetivos, b) aquilo que apreendemos quando entendemos uma expressão linguística. Além disso, Frege afirma que c) sentidos determinam o referente de uma expressão linguística pelo fato deste satisfazer o sentido associado a ela –  é o referente de ‘Superman’ porque  satisfaz o sentido daquele nome. Mas isso é muito vago. Devemos ainda nos perguntar como exatamente sentidos determinam um referente, qual é sua natureza, e como eles são capazes de explicar os Enigmas de Frege.

Com relação ao modo como sentidos determinam um referente, um modo bastante

comum de entender o funcionamento de sentidos é explicá-los em termos de descrições definidas, como fiz quando introduzi a noção de sentido na seção anterior. Uma descrição definida expressa uma condição a ser satisfeita por um e apenas um objeto (se houver tal objeto). No caso de ‘a pessoa que veste cuecas vermelhas por cima da calça’, a condição a ser satisfeita é a de ser uma pessoa que veste cuecas vermelhas por cima da calça. Como apenas  satisfaz essa condição, então a descrição definida determina . Da mesma forma, o sentido de ‘Superman’ expressa uma condição que é satisfeita apenas por . Por isso  é o referente de ‘Superman’.

Se entendermos essa explicação como uma analogia entre o funcionamento de sentidos e descrições definidas, então ela me parece ser bastante consistente com o que Frege afirma em seus textos. A analogia não nos compromete com a controversa afirmação que sentidos são de fato o conteúdo de descrições definidas e, portanto, necessariamente envolvendo conceitos. Pela analogia, o sentido poderia ser como o conteúdo de uma foto: conteúdo que não pode ser expresso usando apenas conceitos, mas que tem condições de satisfação – por exemplo, uma foto minha em frente a uma casa é uma representação de mim (e de uma casa em particular) e não de Bill Gates, apesar de seu conteúdo não poder ser completamente capturado por meio de descrições.

Com relação à natureza de sentidos, Frege argumenta que sentidos são únicos porque não são entidades do *mundo interior*, nem entidades do *mundo exterior*. Para Frege, entidades do mundo interior são aquelas que dependem de alguém “ter elas em mente” para existir. Por exemplo, essa voz interna que o leitor está “ouvindo” enquanto lê este artigo só existe na medida em que o leitor a tem em mente. Se o leitor para de ouvir essa voz interior, ela deixa de existir. O mesmo é o caso de todas as nossas sensações. A dor de cabeça que eu tenho só existe na medida em que eu sinto ela. Se eu deixo de senti-la, ela não continua “voando” por aí até alguém “pegar” ela. Pelo contrário, ela simplesmente deixa de existir. Além do mais, essas entidades mentais não são compartilháveis. Como Frege diz (1956) “Nenhuma outra pessoa sente a minha dor.” A dor de cabeça que tenho agora é minha e apenas minha. As pessoas podem sentir simpatia por mim, mas não podem sentir minha dor.

Já as entidades do mundo exterior são objetos concretos, como carros, árvores, etc. Esses objetos existem independente de alguém percebê-los, e interagem causalmente entre si. Um animal que mora em uma parte inexplorada do fundo mar e que por isso nunca foi visto por ninguém continua existindo. Sua casa continua existindo quando você sai de manhã para trabalhar mesmo que você more no meio da floresta e não tenha ninguém para vê-la enquanto você não a está vendo. Nós não contratamos pessoas para *perceber* nosso carro de noite a fim de ter certeza que ele estará lá no dia seguinte (podemos contratar pessoas para

vigiar os carros e ter certeza que ninguém os roube, mas esse não é um problema metafísico).

Por um lado, afirma Frege, sentidos não podem fazer parte do reino mental. Proposições existem (e possuem valores de verdade) independentemente de serem apreendidas por alguém. Possivelmente existe um número infinito de proposições e muitas delas não serão nunca apreendidas por alguém. Além disso, como expliquei anteriormente, sentidos são compartilháveis e comunicáveis: tanto eu quanto Aline podemos apreender o mesmo sentido quando entendemos o nome ‘Superman’. Por outro lado, para Frege sentidos obviamente não são entidade do mundo exterior; eles não são concretos nem interagem causalmente com carros, ou casas. Portanto, Frege conclui que *sentidos* são *suis generis* e fazem parte de um terceiro reino de objetos, possivelmente o reino dos objetos abstratos junto com números, propriedades, etc.

Alguns dos vários problemas apontados com a conclusão desse argumento é que mesmo sabendo que sentidos fazem parte do terceiro reino, eles continuam sendo misteriosos, e foram postulados *apenas* com a finalidade de ser o conteúdo semântico de expressões linguísticas. Essa afirmação parece ser *ad hoc* a não ser que tenhamos bons argumentos contra as posições alternativas - como que sentidos sejam conceitos ou propriedades - o que não é claro que Frege tenha oferecido⁵. É evidente em seus artigos que ele não gostaria que sentidos fossem conceitos, mas não tão claro se ele oferece boas razões pra isso.

Também não é claro como a essas entidades misteriosas oferecem uma solução para os Enigmas. Justamente por não sabermos o que são sentidos, a explicação de Frege desperta mais perguntas do que responde, enquanto que uma explicação satisfatória do fenômeno deveria fazer o contrário. Explicar um fenômeno *A* recorrendo a *B* sem que tenhamos entendimento prévio de *B* nos faz questionar o que é *B*. Por isso uma explicação envolvendo *B* não é considerada uma explicação satisfatória. No caso de Frege, ele propõe uma explicação do fenômeno presente nos Enigmas recorrendo a noção de sentido. Mas como não temos conhecimento prévio do que é um sentido e Frege também não oferece uma explicação, não é claro que ganhamos uma compreensão do fenômeno por trás dos Enigmas. Sem saber o que são sentidos, não é claro exatamente *como* ‘Superman é Superman.’ e ‘Superman é Clark Kent.’ diferem em valor informativo. Esse problema é bastante embaraçoso para Frege e, se não solucionado, pode realmente por em risco sua teoria.

Um modo bastante comum de responder a essa objeção é identificar sentidos de nomes próprios como o conteúdo descrições definidas, o modo como apresentei a teoria

5 - Um dos motivos que Frege oferece para defender que sentidos não são conceitos é que conceitos podem ser apreendidos por meio de sentidos. Em outras palavras, conceitos possuem modos de apresentação assim como objetos do mundo físico. Mas esse não me parece ser um bom motivo para chegar a essa conclusão porque sentidos também possuem modos de apresentação: mda_{CK}^1 é modo de apresentação de mda_{CK} .

de Frege. Essa interpretação é inspirada no tratamento de nomes próprios proposto por Russell⁶. Teorias que seguem essa linha são chamadas de ‘Teorias Descritivistas’. Mas essa é provavelmente uma interpretação incorreta de Frege, como já mencionado anteriormente e na nota de rodapé 3.

Sob essa interpretação, nomes próprios são, na verdade, abreviações de uma ou mais descrições definidas, que pode ou não variar de falante para falante. *Teorias Descritivistas Simples* sugerem que nomes próprios sejam abreviações de uma descrição definidas. Nesse caso, ‘Superman’ é uma abreviação da descrição ‘a pessoa que veste cuecas vermelhas por cima da calça’. E a sentença ‘Superman voa.’ é uma abreviação de ‘A pessoa que veste cuecas vermelhas por cima da calça voa.’. De forma análoga, o nome ‘Clark Kent’ abrevia a descrição ‘o jornalista do Planeta Diário apaixonado por Lois Lane’, e a sentença ‘Clark Kent voa.’ abrevia a sentença ‘O jornalista do Planeta Diário apaixonado por Lois Lane voa.’.


Sentidos, então, são da mesma natureza que conceitos como o conceito *vermelho*. E apreendemos sentidos da mesma maneira que apreendemos conceitos. Claro, *conceitos* não são exatamente entidades das quais temos conhecimento pleno. Portanto, em certa medida, o problema da explicação satisfatória apresentada alguns parágrafos atrás pode reaparecer. Mas uma vantagem dessa interpretação é que pelo menos já temos uma discussão em andamento sobre o que são conceitos. Então de fato sabemos um pouco mais sobre eles do que sobre *sentidos*.

Com essa interpretação também temos mais claro como sentidos de fato explicam o fenômeno por trás do Enigma. ‘Superman é Superman.’ e ‘Superman é Clark Kent.’ expressam proposições com conceitos diferentes. A primeira diz que «a pessoa que veste cuecas vermelhas por cima da calça é a pessoa que veste cuecas vermelhas por cima da calça». Os mesmos conceitos (organizados da mesma forma) estão envolvidos nos dois lados a identidade. Então essa é uma informação trivial. A segunda diz que «a pessoa que veste cuecas vermelhas por cima da calça é o jornalista do Planeta Diário apaixonado por Lois Lane». Essa informação não é trivial; conceitos diferentes estão envolvidos nos dois lados da identidade.

Da mesma forma temos a explicação para a segunda versão do Enigma. ‘Lois Lane acredita que Superman voa.’ e ‘Lois Lane acredita que Clark Kent voa.’ podem ter referentes diferentes, quer dizer, valores de verdade diferentes porque as expressões que compõem cada sentença se referem a objetos diferentes: ‘Superman’ se refere a mda_{SM} e ‘Clark Kent’ se refere a mda_{CK} .

Teorias descritivistas são geralmente inspiradas na teoria das descrições de Russell


6 - Ver Russell (1910-1911).

(1905, 1910-1911), e o seu maior defensor é muito provavelmente John Searle (1958). A versão de Searle é mais conhecida como ‘Teoria dos Feixes de Descrições’ (*‘Cluster Theory’*) porque propõe que nomes próprios abreviem um conjunto de descrições definidas ao invés de abreviar apenas uma descrição. Para ele ‘Superman’ é associado com um conjunto de descrições, por exemplo, {a pessoa que veste cuecas vermelhas por cima da calça, o super-herói criado por John Siegel, o filho de Jor-El, o habitante do planeta Terra que voa, etc.}, sendo que uma quantidade suficiente dessas descrições, mas não necessariamente todas, é satisfeita pelo seu referente, i.e., .⁷

Uma das vantagens da teoria de Searle com relação à Teoria Descritivista Simples é que ela dá conta de alguns problemas levantados para essa última⁸. Por exemplo, se ‘Superman’ abrevia a descrição ‘a pessoa que veste cuecas vermelhas por cima da calça’, então ‘Superman veste cuecas vermelhas por cima da calça.’ é uma abreviação de ‘A pessoa que veste cuecas vermelhas por cima da calça veste cuecas vermelhas por cima da calça.’ Essa sentença expressa uma verdade analítica; quer dizer, sabendo o significado das palavras envolvidas, sabemos que a sentença é verdadeira - do mesmo modo que ‘Solteiros são não casados.’ é analítica. Entretanto, é absurdo dizer que ‘Superman veste cuecas vermelhas por cima da calça.’ é uma verdade analítica. Uma pessoa pode muito bem saber o que ‘Superman’ significa e não saber se essa sentença é verdadeira ou falsa. Logo, ‘Superman’ não pode abreviar a descrição ‘a pessoa que veste cuecas vermelhas por cima da calça’ e a Teoria Descritivista Simples é falsa.

Na teoria de Searle a sentença ‘Superman veste cuecas vermelhas por cima da calça.’ não é analítica, como não deve ser. Isso porque, grosso modo, ela abrevia a sentença não analítica ‘A pessoa que satisfaz uma quantidade suficiente das descrições: a pessoa que veste cuecas vermelhas por cima da calça, o super-herói criado por John Siegel, o filho de Jor-El, o habitante do planeta Terra que voa, etc. veste cuecas vermelhas por cima da calça.’ Apesar de engenhosa, a teoria de Searle recebeu críticas fortes desde que foi proposta⁹.

Independente dos problemas com a noção de *sentido* proposta por Frege, é inegável que a ideia geral da Teoria Fregeana de recorrer a modos de apresentação para explicar a intuição de que há algo de enigmático nos pares de sentença é bastante atrativa. Mesmo as teorias adversárias, como a Teoria da Referência Direta, usam sentidos enquanto modos


7 - Searle deixa em aberto qual é o número suficiente de descrições que precisam ser satisfeitas para que ‘Superman’ se refira a . Em seu artigo (1958) ele parece sugerir que esse número é determinado contextualmente. Deixo essa discussão de lado porque ela não é pertinente a teorias sobre o conteúdo semântico de nomes próprios, que é o tópico desse artigo, mas sim a teorias que explicam como nomes próprios se referem a objetos. Para mais sobre a distinção ver Kripke (1980)

8 - Para mais objeções, ver Kripke (1980).

9 - Ver Kripke (1980) e Putnam (1975).

de apresentação a fim de salvar-se do problema posto pelos Enigmas. A principal diferença é que, enquanto Frege defende que sentidos fazem parte do conteúdo semântico (das condições de verdade de sentenças), as teorias da referência direta normalmente colocam o sentido em outro tipo de conteúdo que não semântico. Veremos como isso é posto em prática em detalhes na seção 7.

6.2. Cadeia Infinita de Sentidos

Outro problema levantado contra a teoria fregeana é a geração de uma cadeia a princípio infinita de sentidos gerada a partir de iterações de verbos que expressam atitudes proposicionais. Focando apenas no sentido do nome ‘Superman’, na sentença ‘Superman voa.’ ele expressa o sentido mda_{SM} e se refere a . Quando colocamos esta sentença no escopo de um verbo que expressa atitude proposicional, como ‘Lois Lane acredita que Superman voa.’, o nome ‘Superman’ agora expressa um sentido diferente, mda_{SM}^1 e se refere a mda_{SM} . Se colocarmos essa última sentença no escopo de outro verbo que expressa uma atitude proposicional, como em ‘Aline sabe que Lois Lane acredita que Superman voa.’, o nome ‘Superman’ expressa outro sentido mda_{SM}^2 e se refere a mda_{SM}^1 . E assim sucessivamente. Do ponto de vista gramatical, podemos *sempre* adicionar um novo verbo que expressa atitude proposicional a uma sentença. E toda vez que adicionamos um novo verbo, ‘Superman’ expressa um novo sentido. Isso quer dizer que o nome ‘Superman’ expressa um conjunto infinito de sentidos: um sentido ordinário (mda_{SM}), e um sentido para cada verbo que expressa atitude proposicional adicionado a sentença ‘Superman voa.’ (mda_{SM}^1, mda_{SM}^2 , etc.).

À primeira vista, essa cadeia infinita de sentidos por si só não é problemática. O problema surge quando pensamos sobre a relação entre sentidos e o modo como aprendemos uma linguagem.

Linguagens naturais (como português, inglês, espanhol, etc.) são tais que se uma pessoa sabe o significado de ‘Lois Lane’, ‘Superman’, ‘ x voa’ e ‘ x ama y ’, então, em casos normais, essa pessoa também sabe o significado das seguintes sentenças: ‘Superman voa.’, ‘Lois Lane voa.’, ‘Lois Lane ama Superman.’, ‘Lois Lane ama Lois Lane’, ‘Superman ama Superman.’ e ‘Superman ama Lois Lane.’ Uma pessoa que sabe o significado daquelas quatro expressões linguísticas é capaz de entender seis sentenças diferentes. É essencial para explicar nossa capacidade de entender sentenças novas que as linguagens naturais tenham essa propriedade. Se assim não fosse, toda vez que fôssemos apresentados a uma sentença com a qual nunca nos deparamos antes, seria necessário aprender um novo significado. Mas isso

é contra intuitivo. Existe uma boa probabilidade de o leitor nunca ter escutado a sentença ‘Papai Noel é parente do Bill Gates.’ antes. Mesmo assim o leitor é capaz de entender o significado dessa sentença porque já sabe do significado das expressões ‘Papai Noel’ ‘ser parente de’ e ‘Bill Gates’ separadamente. Ou seja, o leitor consegue entendê-la sem precisar aprender nada a respeito do português. Contudo, como vou argumentar no parágrafo seguinte, não parece ser possível que a linguagem natural funcione dessa forma na teoria de Frege, justamente porque de acordo com ela um nome expressa infinitos sentidos.

Na teoria de Frege, entender o nome ‘Superman’ significa saber o sentido de ‘Superman’. Como vimos, o sentido que uma pessoa precisa saber para satisfazer a condição de compreender qualquer sentença em que ‘Superman’ aparece, incluindo sentenças com vários verbos que expressam atitudes proposicionais, é um conjunto infinito de sentidos. Mas é claramente implausível supor que ao compreender o nome ‘Superman’, um falante apreende um número infinito de sentidos. Isso não só é psicologicamente impossível, dada que a capacidade de compreensão humana é finita, como também aparentemente desnecessário. Por causa disso, talvez a teoria de Frege não seja apropriada para explicar linguagens naturais e, conseqüentemente, para explicar os dois Enigmas de Frege.

Uma possível resposta a essa objeção é dizer que não é necessário saber todos os infinitos sentidos que ‘Superman’ pode expressar. Basta saber o sentido expresso em contextos ordinários que a partir dele de alguma forma inferimos o sentido em contextos indiretos. Por exemplo, sabemos que o sentido ordinário de ‘Superman’ é mda_{sm} . Quando então o nome aparece no escopo de um verbo que expressa atitude proposicional, como em ‘Lois Lane acredita que Superman voa.’, inferimos que seu sentido é mda_{sm}^1 porque sabemos de antemão que o seu referente nessa sentença é mda_{sm} . E o mesmo no caso de ‘Aline sabe que Lois Lane acredita que Superman voa.’ Porque temos dois verbos que expressam atitudes proposicionais, sabemos que o sentido de ‘Superman’ (mda_{sm}^2) vai se referir ao sentido (mda_{sm}^1) que se refere ao sentido ordinário (mda_{sm}), sendo que este último sabemos porque compreendermos o nome ‘Superman’ em contextos ordinários. Nessa proposta, apenas sabendo o sentido ordinário de ‘Superman’ somos capazes de entender qualquer sentença em que o nome aparece, dado que também saibamos o sentido das outras expressões que compõe a sentença.

Essa sugestão parece de fato dar conta do problema em questão. Mas infelizmente ela não está disponível a Frege. Frege diz explicitamente que enquanto é possível determinar o referente de uma expressão a partir do sentido expresso por ela, o contrário não é possível. Ou seja, não temos como saber o sentido de uma expressão apenas sabendo qual é seu referente, que é o que a sugestão apresentada aqui exige. A razão para Frege dizer isso é simples. Cada sentido determina um e apenas um objeto (se houver algum). Em outras

palavras, cada sentido é modo de apresentação de um a apenas um objeto (se houver algum). Por isso com o sentido podemos determinar o referente. Mas para cada objeto há infinitos sentidos associados a ele – como no caso de \uparrow que pode estar junto tanto com mda_{sm} quanto com mda_{CK} . Então o referente de uma expressão não é suficiente para determinar seu sentido; saber que \uparrow é o referente de ‘Superman’ não nos dá informação suficiente para descobrir quando é seu o sentido, se mda_{sm} , mda_{CK} ou outro modo de apresentação de \uparrow .

Alguns filósofos são céticos a respeito da força dessa objeção. De acordo com eles temos uma sentença no escopo de vários verbos que expressam atitudes proposicionais, não é mais muito claro qual é o significado dessa sentença ou suas condições de verdade. E se é o sentido que fornece as condições de verdade, então não é claro que realmente sabemos o sentido das expressões que compõem a sentença. Por exemplo, não é muito claro que sabemos o que está sendo dito ou que condições devem ser satisfeitas para a seguinte sentença ser verdadeira: ‘Juliana disse que Bill Gates suspeita que Aline sabe que Lois Lane acredita que Superman voa.’ De acordo com esses filósofos céticos, isso quer dizer que talvez nós de fato não sabemos qual é o sentido de ‘Superman’, dentre outros. Então quando anteriormente dissemos que sabemos o sentido de ‘Superman’ e que somos capazes de entender quaisquer sentenças em que o nome aparece, na verdade queríamos dizer que sabemos seu sentido ordinário e seu sentido em contextos indiretos quando no escopo de apenas um ou no máximo dois verbos que expressam atitudes proposicionais.

Essa linha de resposta não me parece muito atrativa. Concordo que à primeira vista é difícil acompanhar o que está sendo dito na sentença acima. Mas existem outras explicações mais plausíveis para nossa falta de compreensão de sentenças daquele tipo. Uma delas é que não estamos acostumados a proferir, ouvir ou ler sentenças com mais de um ou dois verbos que expressam atitudes proposicionais. Por causa disso temos dificuldades em acompanhar o significado desse tipo de sentenças. Essa dificuldade é bastante similar com a dificuldade que temos de acompanhar o significado de sentenças longas (lembre-se de quando o leitor lê Kant!), e não está relacionada com a falta de conhecimento dos sentidos que compõem elas. Portanto, a dificuldade de entender o significado da sentença acima seria puramente psicológica, relacionada com nossa capacidade de manter na mente ao mesmo tempo uma quantidade limitada de informações. Essa linha de explicação me parece ser muito mais intuitiva que aquela que diz que não sabemos mais o significado de ‘Superman voa.’ na sentença longa.

Outro modo bastante intuitivo de tentar salvar a teoria de Frege dessa objeção é dizer que para toda expressão há apenas dois sentidos: o sentido em contextos indiretos e o sentido em contextos diretos. O sentido em contextos diretos é o sentido ordinário

como Frege sustentou, como o sentido mda_{sm} expresso por ‘Superman’ em ‘Superman voa.’. A novidade da proposta é afirmar que para cada expressão linguística só há um sentido expresso em contextos indiretos (ao invés de infinitos sentidos), não importando quantos verbos que expressam atitudes proposicionais estão presentes na sentença. Nesse caso, o nome ‘Superman’ nas sentenças ‘Lois Lane acredita que Superman voa.’ e ‘Aline sabe que Lois Lane acredita que Superman voa.’ expressam exatamente o mesmo sentido mda_{sm}^1 , que se refere ao sentido ordinário mda_{sm} .

Como nessa sugestão uma expressão linguística pode expressar *apenas dois sentidos*, ao invés de um conjunto infinito de sentidos, o problema explicado no começo da seção não aparece. Também não é implausível ou psicologicamente impossível afirmar que quando compreendemos o nome ‘Superman’, apreendemos dois sentidos: o sentido ordinário e o sentido em contextos indiretos.

Essa solução pode parecer ser satisfatória à primeira vista. Mas ela também enfrenta problemas sérios que serão apresentados na próxima seção.

6.3. Sentido de quem, afinal de contas?

O último problema que vou discutir está relacionado com a afirmação de Frege que o sentido associado a uma expressão linguística pode variar de pessoa pra pessoa. Suponha que Lois Lane associe apenas o sentido mda_{sm} - «a pessoa que veste cuecas vermelhas por cima da calça» - a ‘Superman’, e que Aline associe apenas o sentido «o filho de Jor-El» - que vou abreviar como ‘ mda_{sm}^* ’. Como o nome ‘Superman’ em ‘Aline sabe que Lois Lane acredita que Superman voa.’ está em um contexto indireto, sabemos que ele expressa um sentido, a saber, mda_{sm}^1 , e se refere a outro sentido. A pergunta agora é: qual é esse sentido que é referente de ‘Superman’ - e portanto relevante para o valor de verdade da sentença - e do qual mda_{sm}^1 é um modo de apresentação? O sentido que Lois Lane apreende (mda_{sm}) ou o sentido que Aline apreende (mda_{sm}^*)?

Mark Richard (1990, pp. 69-75) argumenta que nenhuma das duas opções são viáveis, o que quer dizer que a teoria de Frege tem dificuldades em explicar justamente aqueles tipos de sentenças que pareciam dar suporte a ela.

Primeiro, considere a sugestão que o referente seja mda_{sm} , o sentido que Lois Lane associa a ‘Superman’. Esse sentido também é o sentido ordinário de ‘a pessoa que veste cuecas vermelhas por cima da calça’. Ou seja, ‘Superman’ e ‘a pessoa que veste cuecas vermelhas

por cima da calça' têm o mesmo sentido ordinário. O que quer dizer que os seus sentidos em contextos indiretos determinam o mesmo referente, a saber, mda_{sm} . Isso tem como consequência, dentre outras coisas, que se 'Aline sabe que Lois Lane acredita que **Superman** voa.' é verdadeira, então 'Aline sabe que Lois Lane acredita que **a pessoa que veste cuecas vermelhas por cima da calça** voa.' também é verdadeira¹⁰. Porém, Richard argumenta que essa implicação é falsa porque é possível que a primeira sentença seja verdadeira e a segunda falsa. Suponha que Aline ouça Lois Lane sinceramente afirmando que ela (Lois Lane) acredita que Superman voa, mas que Aline não associe o sentido mda_{sm} ao nome 'Superman' e também não saiba que Superman é a pessoa que veste cuecas vermelhas por cima da calça (porque ela nunca viu Superman). Nesse caso, parece intuitivo que Aline sabe que Lois Lane acredita que **Superman** voa, mas falso que Aline sabe que Lois Lane acredita que **a pessoa que veste cuecas vermelhas por cima da calça** voa – tornando a implicação anterior falsa. Portanto, a sugestão que o referente de 'Superman' em contextos indiretos é mda_{sm} , ou o sentido de Lois Lane deve ser descartada.

A segunda proposta é que mda_{sm}^1 determina mda_{sm}^* , o sentido que Aline associa com 'Superman'. Mas agora parece que a teoria implica que 'Aline sabe que Lois Lane acredita que Superman voa.' é falsa quando sabemos que é verdadeira! Sob essa sugestão, o que precisa ser o caso agora para ela ser verdadeira é que Aline sabe que Lois Lane acredita que *o filho de Jor-El* voa. Mas isso é falso. Por suposição o único sentido que Lois Lane associa ao nome 'Superman' é «a pessoa que veste cuecas vermelhas por cima da calça» e não «o filho de Jor-El». Portanto, Lois Lane acredita que *a pessoa que veste cuecas vermelhas por cima da calça* voa mas não que *o filho de Jor-El* voa.

Com base no argumento de Richard que acabamos de ver, a Teoria Fregeana não explica como a sentença 'Aline sabe que Lois Lane acredita que Superman voa.' pode ser verdadeira porque não é claro qual é o referente de 'Superman'. Portanto, a teoria deve ser abandonada.

A Teoria Fregeana do Significado começou bem, com soluções atrativas para os enigmas. Entretanto, ela parece ter sérios problemas quando examinada com mais cuidado. O primeiro é que não é claro o que é um sentido e que, portanto, não é claro qual é a explicação dos Enigmas. Em seguida vimos que a teoria gera um a cadeia infinita de sentidos que preveniria alguém de saber o sentido de expressões linguísticas. Por fim, vimos o problema de decidir qual é o referente de 'Superman' em 'Aline sabe que Lois Lane acredita que Superman voa.': o sentido associado por Aline ou aquele associado por Lois Lane?

Não é muito claro como Frege poderia responder satisfatoriamente às objeções

¹⁰ - Essas duas sentenças podem expressar proposições diferentes mas como todos os seus componentes tem o mesmo referente, elas se referirão ao mesmo valor de verdade.

levantadas aqui. Então talvez seja uma boa ideia abandonar momentaneamente a teoria de Frege e nos voltarmos pra teoria da referência direta. Mas mesmo que o leitor discorde que esses são problemas graves à teoria, ainda assim é interessante saber como a teoria da referência direta responde ao problema levantado pelos dois enigmas. Na próxima seção vou apresentar a proposta de Salmon de explicar os enigmas sem rejeitar a Teoria da Referência Direta.

7. Teoria da Referência Direta Revisitada

Antes de explicar a teoria proposta por Salmon, gostaria de recapitular rapidamente os dois enigmas que levantam objeções à teoria da referência direta. O primeiro Enigma é explicar como as sentenças ‘Lois Lane acredita que Superman voa.’ e ‘Lois Lane acredita que Clark Kent voa.’ têm valores de verdade diferentes. Para a Teoria da Referência Direta elas expressam a mesma proposição e, portanto, deveriam ter o mesmo valor de verdade. O segundo Enigma é explicar a diferença com relação à informatividade e modo de conhecimento das proposições expressas por ‘Superman é Superman.’ e ‘Superman é Clark Kent.’ Já que para a Teoria da Referência Direta essas duas sentenças expressam o mesmo conteúdo semântico, elas deveriam ser ou ambas informativas, ou ambas não informativas, o que contradiz nossas intuições.

Em geral, a estratégia usada pelos defensores da Teoria da Referência Direta para explicar os enigmas é dizer que quando avaliamos a informatividade de sentenças ou as crenças expressas por elas, nós levamos em consideração não apenas o conteúdo semântico da sentença mas também outros conteúdos ou informações associadas às sentenças. E enquanto o conteúdo semântico de sentenças com nomes diferentes co-referenciamas de outra forma idênticas é o mesmo, os conteúdos não semânticos associados a elas são diferentes. É essa diferença nos conteúdos não semânticos que explica a diferença na informatividade e a aparente diferença no valor de verdade das sentenças que reportam o conteúdo da crença de Lois Lane.

Aqui eu vou discutir a teoria de Salmon, uma das várias que seguem a linha de resposta esboçada no parágrafo anterior. Vou focar nela por causa da sua universalidade: Salmon claramente apresenta sua proposta como uma que elucida o Enigma de Frege envolvendo não só nomes próprios (as versões discutidas aqui) mas também outras versões que citei mas deixei de lado. Quer dizer, diferente de outras soluções, a proposta de Salmon tem a vantagem de unificar as respostas às várias versões do Enigma de Frege. Isso captura a

intuição de que as diferentes versões do Enigma são de fato versões de um só fenômeno.

7.1. Proposta do Salmon

Salmon defende uma versão da Teoria da Referência Direta, portanto ele defende a Tese 1 e 2. Consequentemente, ele se compromete com a Tese 3 e 4. Em particular, ele é enfático na sua posição de que tanto ‘Lois Lane acredita que Superman voa.’ quanto ‘Lois Lane acredita que Clark Kent voa.’ são verdadeiras. Mas e nossas intuições de que a primeira, mas não a segunda é verdadeira? Em um certo sentido, ele rejeita que nossas intuições sejam confiáveis. Em particular, Salmon rejeita que nossas intuições sobre o valor de verdade daquelas sentenças sejam sobre o valor de verdade do seu conteúdo semântico, e argumenta que elas são sobre um outro conteúdo associado às sentenças. E sem haver diferença entre valores de verdade das duas sentenças, o Enigma deixa de existir.

Relembremos que sentenças expressam conteúdos semânticos e conteúdos não semânticos. Quando respondo a pergunta da mãe de Aline a respeito de sua nacionalidade proferindo ‘Aline é brasileira.’ com uma certa expressão facial, eu expresso o conteúdo semântico que informa que Aline é brasileira. Com isso também comunico pragmaticamente a informação que suspeito de algo estranho com a mãe de Aline. Esse conteúdo também tem condições de verdade. No exemplo aqui pode ser verdadeiro ou falso que eu suspeito de algo errado com a mãe de Aline. E como Salmon observa, os valores de verdade do conteúdo semântico e do conteúdo pragmático são independentes. Quer dizer, é possível que o conteúdo semântico seja verdadeiro, mas o conteúdo pragmático falso - isso seria o caso se Aline fosse de fato brasileira mas eu não suspeitasse que tem algo de errado com sua mãe.

O mesmo raciocínio se aplica com as sentenças ‘Lois Lane acredita que Superman voa.’ e ‘Lois Lane acredita que Clark Kent voa.’, de acordo com Salmon. Ambas expressam um conteúdo semântico e comunicam um conteúdo pragmático. Como Salmon é defensor da Teoria da Referência direta, o conteúdo semântico que elas expressam é necessariamente o mesmo: $\langle \text{Lois Lane, acreditar, } \langle \text{ } \rangle, \text{ voa} \rangle$. O conteúdo pragmático, por outro lado, pode ser diferente. E nesse caso é diferente. Para Salmon, a primeira sentença expressa um conteúdo pragmático que pode ser mais ou menos do tipo: *Lois Lane acredita na proposição expressa por ‘Superman voa.’ através do modo de apresentação que ela associa à ‘Superman voa.’.* Já a segunda sentença expressaria o seguinte conteúdo pragmático: *Lois Lane acredita na proposição expressa por ‘Clark Kent voa.’ através do modo de apresentação que ela associa à ‘Clark Kent voa.’.*

Para Salmon, modos de apresentação de proposições pragmaticamente comunicados por sentenças são bastante semelhantes ao sentido fregeanos, modos de apresentação de objetos. Podem existir diferentes modos de apresentação para cada proposição, assim como podemos ter diferentes modos de apresentação para cada referente. E assim como diferentes modos de apresentação de um objeto podem ser expressos por diferentes nomes, diferentes modos de apresentação de proposições podem ser pragmaticamente comunicados por diferentes sentenças. Por exemplo, a proposição $\langle \lambda, \text{voa} \rangle$ pode ser associada a um modo de apresentação pragmaticamente comunicado pela sentença ‘Clark Kent voa.’ e a outro modo de apresentação diferente comunicado pela sentença ‘Superman voa.’ – tal como na teoria de Frege λ pode ser associado a mda_{CK} que é expresso por ‘Clark Kent’, ou associado a mda_{SM} que é expresso por ‘Superman.’ A diferença entre as duas teorias é que enquanto Frege considera o modo de apresentação parte do conteúdo semântico expresso por uma expressão linguística, Salmon considera o modo de apresentação parte do conteúdo pragmaticamente comunicado.

Com base no que sabemos sobre Lois Lane, é evidente que o conteúdo pragmaticamente comunicado por ‘Lois Lane acredita que Superman voa.’ é verdadeiro e o conteúdo comunicado por ‘Lois Lane acredita que Clark Kent voa.’ é falso. O que precisa ser explicado agora é como é possível que o conteúdo semântico tenha o mesmo valor de verdade - nesse caso ambas são verdadeiras -, como a Teoria da Referência Direta antecipa?

Para responder a essa pergunta, Salmon sugere que uma pessoa acredita em uma proposição se, e somente se, ela estiver em uma relação acreditar*¹¹ (que é diferente de *acreditar*) na proposição através de pelo menos um modo de apresentação. No caso de Lois Lane, ela acredita em $\langle \lambda, \text{voa} \rangle$ se, e somente se, ela acreditar* nessa proposição através de pelo menos um modo de apresentação. Essa última afirmação é verdadeira de acordo com a estória: Lois Lane acredita em $\langle \lambda, \text{voa} \rangle$ através do modo de apresentação associado e pragmaticamente comunicado por ‘Superman voa.’ Como de acordo com a referência direta a sentença ‘Lois Lane acredita que Clark Kent voa.’ é verdadeira se, e somente se, Lois Lane acreditar na proposição expressa por ‘Clark Kent voa.’, pela conclusão anterior se segue que ela é verdadeira¹².

Resumidamente, Salmon propõe que o conteúdo semântico de ‘Lois Lane acredita que Superman voa.’ e de ‘Lois Lane acredita que Clark Kent voa.’ são o mesmo, e ambas são verdadeiras. Entretanto, conteúdo pragmático comunicado pelas duas sentenças são

11 - ‘acreditar*’ se refere à mesma relação que Salmon chama de ‘BEL’.

12 - Caso o leitor esteja curioso, para Salmon, Lois Lane não acredita em $\langle \lambda, \text{voa} \rangle$ se, e somente se, não é o caso que ela acredite* em $\langle \lambda, \text{voa} \rangle$ através de pelo menos um modo de apresentação. Em outras palavras, Lois Lane não acredita em $\langle \lambda, \text{voa} \rangle$ se, e somente se, não existir um modo de apresentação de λ através do qual ela acredita que $\langle \lambda, \text{voa} \rangle$.

diferentes; a primeira comunica um conteúdo verdadeiro e a segunda um conteúdo falso. Mas o valor de verdade do conteúdo pragmático não está relacionado com o valor de verdade do conteúdo semântico.

A proposta de Salmon faz com que os Enigmas de Frege desapareçam porque rejeita uma de suas suposições. O primeiro Enigma era: como é possível as sentenças ‘Lois Lane acredita que Superman voa.’ e ‘Lois Lane acredita que Clark Kent voa.’ terem diferentes valores de verdade sendo que ‘Superman’ e ‘Clark Kent’ são co-referenciais? Na teoria de Salmon, essas duas sentenças não diferem em valor de verdade, então não há enigma a ser explicado. E nossa intuição de que os valores de verdade das sentenças são diferentes? Para Salmon essa intuição é sobre o valor de verdade do conteúdo pragmaticamente comunicado e não do conteúdo semântico, como erroneamente foi entendido por Frege e outros filósofos. As duas sentenças de fato diferem com relação ao valor de verdade do conteúdo pragmaticamente comunicado. Mas essa diferença não gera nenhum problema para a Teoria da Referência Direta, muito menos sugere que a Tese 2 é falsa. A Teoria da Referência Direta é uma teoria do conteúdo semântico e não do conteúdo pragmático. A diferença no valor de verdade das sentenças é com relação ao conteúdo pragmático. Portanto se essa diferença gera um problema, ela gera um problema para teorias sobre o conteúdo pragmático e não para Teorias da Referência Direta.

Para o segundo enigma, a solução de Salmon segue a mesma estratégia geral da solução ao primeiro enigma. O enigma era como é possível que ‘Superman é Superman.’ e ‘Superman é Clark Kent.’ tenham diferentes propriedades epistêmicas (a primeira é conhecível *a priori* enquanto a segunda não) e valor informativo (a primeira não é informativa enquanto a segunda é)? Novamente, para o Enigma surgir, temos que assumir que nossas intuições a respeito de suas propriedades epistêmicas e seus valores informativo são sobre o conteúdo semântico expresso por elas. Para Salmon isso é um erro que cometemos porque essas intuições são sobre o conteúdo pragmaticamente comunicado. Apesar dos conteúdos semanticamente expressos por ‘Superman é Superman.’ e por ‘Superman é Clark Kent.’ serem o mesmo, os conteúdos pragmaticamente comunicados não o são. O conteúdo pragmático comunicado pela primeira poderia ser algo como: o referente de ‘Superman’ é idêntico ao referente de ‘Superman.’ E aquele comunicado pela segunda sentença pode ser algo como o referente de ‘Superman’ é idêntico ao referente de ‘Clark Kent.’ O primeiro é obviamente não informativo e conhecível *a priori*. O segundo, por outro lado, é evidentemente informativo e não é conhecível *a priori*. Dessa forma explicamos nossas intuições de que tem alguma coisa de diferente entre as sentenças sem precisar dizer que seus conteúdos semânticos sejam diferentes, ou seja, sem rejeitar a Teoria da Referência Direta.

Para concluir, de acordo Salmon, quando fazemos a distinção entre o conteúdo semanticamente expresso e o conteúdo pragmaticamente comunicado percebemos que o problema posto pelos Enigmas de Frege para a Teoria da Referência Direta desaparecem. Nossas intuições a respeito da diferença entre informações associadas com as sentenças são sobre o conteúdo pragmaticamente comunicado e não sobre o conteúdo semanticamente expresso.

8. Problemas para a solução do Salmon

A teoria do Salmon me parece ser bastante engenhosa. Ela lida aparentemente bem com as duas versões do Enigma de Frege com nomes próprios discutidas aqui, além de oferecer uma resposta homogênea para outras versões recentes do enigma. Então por que não abraçar essa teoria e dar o problema gerado pelo Enigma de Frege como solucionado? Infelizmente a teoria do Salmon me parece ter problemas sérios que serão apresentados nas próximas seções.

A solução de Salmon consiste em argumentar que o Enigma de Frege não causa problema para a Teoria da Referência Direta porque: a) existe um conteúdo não semântico associado com sentenças, o conteúdo pragmaticamente comunicado, b) que é normalmente confundido com o conteúdo semântico e c) é fonte das nossas intuições com relação aos diferentes valores de verdade das sentenças que reportam as crenças de Lois Lane. Vou apresentar três objeções aqui, uma para cada parte da explicação.

8.1. O que é o conteúdo pragmaticamente comunicado?

A primeira parte da explicação é afirmar a existência de um conteúdo pragmaticamente comunicado por sentenças. Mas que conteúdo é esse? Em nenhum de seus artigos Salmon esclarece o que seria esse conteúdo pragmaticamente comunicado por sentenças. O máximo que ele diz a respeito do conteúdo é que ele *pode* ser (mas não necessariamente é) como o sentido de Frege. Em conversas particulares, Salmon esclarece que não se compromete e não quer se comprometer com uma teoria a respeito do conteúdo pragmaticamente comunicado porque acredita que esse tópico não faz parte do escopo de teorias semânticas. Essas lidam, obviamente, com a natureza do conteúdo semântico e não de conteúdos pragmaticamente comunicados. Portanto, a pergunta sobre o que é esse conteúdo pragmaticamente comunicado não está no escopo da sua teoria ou de qualquer outra teoria semântica.

A resposta de Salmon me parece bastante insatisfatória. Para explicar porque ela me parece insatisfatória, vou apresentar um caso análogo onde uma explicação com a mesma estrutura da de Salmon é intuitivamente insatisfatória. Por analogia, então, inferimos que a explicação original é também insatisfatória.

Considere o seguinte cenário: eu sou uma bióloga que investigo a razão de haver similaridades e diferenças entre animais. Enquanto bióloga, eu lido apenas com animais e não com problemas metafísicos, como por exemplo se há propriedades ou não, ou até com problemas a respeito da existência de Deus. Depois de ter feito várias pesquisas, proponho que a razão de haver similaridades e diferenças entre animais é que Deus criou todas as criaturas e fez algumas bem diferentes e outras bastante similares. Mas eu não incluo nenhuma explicação de Deus ou como ele cria animais.

A reação racional esperada dos meus colegas biólogos perante minha explicação parece ser a de exigir um esclarecimento do que é Deus e como ele cria animais. E se por acaso eu me negar a oferecer tal esclarecimento, então minha teoria deveria ser rejeitada. Agora suponha que respondo de maneira similar a Salmon: como sou uma bióloga e não uma filósofa, explicações a respeito de Deus ou de e como ele age estão fora do escopo da minha pesquisa, portanto não preciso oferecer uma. Parece ser relativamente claro aqui que essa resposta não é suficiente. Não podemos postular entidades em campos de pesquisa que não são os nossos para solucionar um problema no nosso campo de pesquisa e não oferecer qualquer explicação. Por analogia, o mesmo pode ser dito a respeito da proposta do Salmon. Até que ele ofereça uma explicação do que é esse conteúdo pragmaticamente comunicado, como ele é comunicado por usos de sentenças e como sentenças que diferem apenas com relação a nomes co-referenciais podem ter valores de verdade, informativos e modo de conhecimento diferentes, sua proposta deve ser rejeitada como uma solução apropriada para o problema posto pelo enigma de Frege.

8.2. Engano massivo

Outro problema da solução de Salmon parece ser o fato de ela depender que todos nós estamos massivamente enganados a respeito do conteúdo de nossas crenças, desejos, intenções, etc.

O fenômeno ilustrado pelos pares de sentenças envolvendo crenças discutidos aqui é generalizado. Quer dizer, é possível gerar pares de sentenças que reportam atitudes proposicionais com diferentes valores de verdade de virtualmente todas as atitudes

proposicionais de um sujeito. E Salmon espera que sua solução no caso de sentenças que reportam crenças seja também aplicada em caso de sentenças que reportam outras atitudes proposicionais. Ou seja, não só estamos enganados a respeito do conteúdo de nossas crenças, como também do conteúdo de desejos, intenções, etc. Para ele, Lois Lane está enganada quando ela diz que não deseja abraçar Clark Kent. Ela tem esse desejo simplesmente porque deseja abraçar Superman¹³.

O problema dessa consequência é que parece ser bastante plausível assumir que se houvesse um engano massivo a respeito do valor de verdade de sentenças que reportam crenças, desejos, intenções, etc., então ele já teria aparecido. Cotidianamente atribuímos crenças (e outras atitudes proposicionais) a pessoas para explicar comportamento das pessoas. Por exemplo, quando Lois Lane pula no colo de Superman quando está no topo de um prédio em chamas, justificamos seu comportamento dizendo que Lois Lane acredita que Superman voa. Fazemos isso usando sentenças que reportam crenças, como ‘Lois Lane acredita que Superman voa.’ E nós temos intuições sobre quais sentenças são adequadas para explicar um comportamento e quais não são. No caso de Lois Lane, ‘Lois Lane acredita que Clark Kent voa.’ não aparenta ser adequada para explicar seu comportamento – nesse caso porque julgamos ser falsa. Além do mais, nós agimos com base no que intuitivamente julgamos ser o conteúdo de sentenças que reportam crenças e seus valores de verdade. Por exemplo, sabendo que ‘Lois Lane acredita que Superman voa.’ pode motivar Lois Lane a pular no colo de Superman e se salvar, eu vou dizer a ela que Superman está atrás dela e não que Clark Kent está atrás dela.

Parece que se de fato estivéssemos massivamente enganados a respeito do conteúdo semântico e valor de verdade de sentenças que reportam crenças, como Salmon afirma que estamos, então esse engano já teria aparecido em situações cotidianas. Mas curiosamente nosso engano só aparece quando analisamos pares específicos de sentenças que reportam crenças, nomeadamente aqueles que trazem à tona o Enigma de Frege, e assumimos a Teoria da Referência Direta. Isso leva a crer que o problema está com a Teoria da Referência Direta e não com nossas intuições.

13 - Para Salmon, ‘Lois Lane não deseja abraçar Clark Kent.’ é falsa pelo mesmo motivo que ‘Lois Lane não acredita que Clark Kent voa.’ é falsa. Lois Lane não deseja abraçar λ se, e somente se, não é o caso que Lois Lane deseja* abraçar λ através de pelo menos um modo de apresentação. Em outras palavras, Lois Lane não deseja abraçar λ se, e somente se, não há um modo de apresentação de λ através do qual ela deseje abraçar λ . Portanto, Lois Lane não tem desejos inconsistentes. Agradeço Sagid Salles por sugerir uma explicação mais detalhada desse ponto.

8.3. “Empurrando o problema com a barriga”

Os Enigmas de Frege como originalmente propostos estão relacionados com o conteúdo em que encontramos o valor informativo e o conteúdo de crenças que influenciam nossos comportamentos – este último explicado na seção anterior e que vou chamar de ‘valor cognitivo’¹⁴. Claramente sentenças comunicam esse conteúdo. Portanto deve haver uma relação sólida entre o conteúdo com valor informativo e cognitivo, e o conteúdo semanticamente expresso por sentenças.

Frege tenta acomodar nossas intuições sobre essa relação entre o conteúdo semântico e o conteúdo com valor informativo e cognitivo sugerindo que eles sejam apenas um único conteúdo. Por isso, para Frege os Enigmas aparecem para teorias semânticas. Mas os Enigmas são fundamentalmente sobre o conteúdo com valor informativo e cognitivo associado a sentenças.

Salmon rejeita a suposição de Frege que o conteúdo com valor informativo e cognitivo é o conteúdo semanticamente expresso. Ao invés disso, ele defende que o primeiro conteúdo é idêntico ao conteúdo pragmaticamente comunicado. Portanto, se os Enigmas de Frege existem e são pertinentes ao conteúdo com valor informativo e cognitivo, então na proposta de Salmon eles devem aparecer no conteúdo pragmaticamente comunicado. Nesse caso, é inadequado tratar a teoria de Salmon como uma *solução* para os Enigmas de Frege. Primeiro porque ela não oferece uma explicação adequada de como o conteúdo pragmaticamente comunicado por sentenças que expressam o mesmo conteúdo semântico podem diferir (ver seção 8.1). Segundo, porque a teoria de Salmon não elimina o Enigma. Ela se torna apenas uma proposta em que esclarece que o enigma não é um enigma para teorias semânticas mas “empurra o problema com a barriga” para teorias que lidam com o conteúdo pragmaticamente comunicado. Porque agora o enigma é: como é possível que o conteúdo pragmaticamente comunicado pelas sentenças ‘Superman é Superman.’ e ‘Superman é Clark Kent.’ sejam diferentes se os nomes ‘Superman’ e ‘Clark Kent’ são co-referenciais?

Agora também fica mais claro porque Salmon não pode sugerir que o conteúdo pragmaticamente comunicado pode ser sobre a linguagem, como expliquei no fim da seção 7.1. A sugestão de Salmon era de que o conteúdo pragmaticamente comunicado por ‘Superman é Superman.’ pode ser algo como *o referente de ‘Superman’ é idêntico ao referente de ‘Superman’*; enquanto que aquele comunicado por ‘Superman é Clark Kent.’ pode ser

14 - Minha terminologia aqui difere radicalmente da de Salmon e é importante manter isso em mente, caso contrário minha discussão nessa seção pode parecer absurda. Salmon usa ‘valor cognitivo’ or ‘valor informativo’ quase como sinônimo de ‘proposição’. Esse uso é completamente diferente do meu uso de ‘valor cognitivo’ aqui. ‘Valor cognitivo’ como está sendo usado aqui se refere àquele conteúdo que intuitivamente explica nosso comportamento.

algo como o referente de ‘*Superman*’ é idêntico ao referente de ‘*Clark Kent*’. Isso porque, como o próprio Frege (1892) argumenta no começo de seu artigo, o valor informativo ou cognitivo de sentenças não parece ser sobre a linguagem e sim sobre o mundo. Saber que o referente de ‘*Superman*’ é idêntico ao referente de ‘*Clark Kent*’ é um conhecimento linguístico. Mas saber que Superman é Clark Kent, é um conhecimento sobre os objetos no mundo.

9. Considerações finais

Obviamente, não tivemos tempo de analisar todas as propostas para solucionar os Enigmas de Frege. Não é claro que hoje em dia há uma boa explicação para o fenômeno. Existem inúmeras propostas que tentam salvar tanto a Teoria Fregeana do Significado (ou teorias inspiradas na teoria de Frege) quanto a Teoria da Referência Direta. Mas elas acabam me parecendo insatisfatórias por algum motivo ou outro – geralmente esses motivos são similares as considerações apresentadas nesse artigo.

Aqui argumentei que a Teoria Fregeana do Significado e a Teoria da Referência Direta não oferecem uma solução apropriada para os Enigmas de Frege. A Teoria Fregeana é insatisfatória porque aparentemente não há uma explicação detalhada do que seria um sentido (6.1). Além do mais, a teoria nos compromete com a tese de que entender um nome próprio significa apreender um conjunto infinito de sentidos (6.2), o que é psicologicamente impossível. Por fim, como Frege permite que o sentido de um nome varie de pessoa para pessoa, quando temos sentenças que reportam crenças que alguém tem da crença de outra pessoa (‘Aline sabe que Lois Lane acredita que Superman voa.’), não é claro qual sentido é relevante e quais sejam suas condições de verdade (6.3).

A Teoria da Referência Direta Revisitada é bastante atraente à primeira vista, mas também parece ter problemas sérios. Argumentei que a teoria de Salmon não explica os Enigmas de Frege porque recorre a conteúdos que não são explicados, a saber, o conteúdo pragmaticamente comunicado (8.1). Ademais, a teoria exige que tratemos como não confiáveis intuições fundamentais, sem oferecer nenhuma razão a não ser a vontade de sustentar a Teoria da Referência Direta como a teoria semântica (8.2). Por fim, mesmo que aceitemos a proposta de Salmon, os Enigmas continuam sem solução porque passam a ser enigmas para o conteúdo pragmaticamente comunicado (8.3).

Referências

- DUMMETT, M. (1981). *Frege: Philosophy of Language* (2nd ed.). Cambridge, MA: Harvard University Press.
- DUMMETT, M. (1991). *Frege and Other Philosophers*. Oxford: Oxford University Press.
- EVANS, G. (1982). *The Varieties of Reference*. (J. McDowell, Ed.) Oxford New York: Clarendon Press Oxford University Press.
- FREGE, G. (1956). The Thought: A Logica Inquiry. *Mind*, 65(259), 289-311.
- FREGE, G. (1994). On Sense and Reference (1892). In R. M. Harnish (Ed.), *Basic Topics in the Philosophy of Language* (M. Black, & P. Geach, Trans., pp. 142-160). Upper Saddle River, N.J: Prentice-Hall.
- KAPLAN, D. (1968). Quantifying In. *Synthese*, 19(1/2), 178-214.
- KRIPKE, S. A. (1980). *Naming and Necessity*. Oxford: Blackwell.
- PUTNAM, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*, 7, 131-193.
- RICHARD, M. (1990). *Propositional Attitudes: an Essay on Thoughts and How we Ascribe Them*. Cambridge: Cambridge University Press.
- RUSSELL, B. (1905). On Denoting. *Mind*, 14(56), 479-493.
- RUSSELL, B. (1910-1911). Knowledge by Acquaintance and Knowledge by Description. *Proceedings of the Aristotelian Society, New Series*, 11, 108-128.
- SEARLE, J. R. (1958). Proper Names. *Mind*, 67(266), 166-173.

Uma breve introdução ao Paradoxo da Conhecibilidade

A brief introduction to the Knowability Paradox

Iago Bozza Francisco
University of Miami

Resumo

O assunto deste artigo é o paradoxo da conhecibilidade. Esse paradoxo é uma demonstração lógica da proposição de que se todas as verdades são conhecíveis, então todas as verdades são conhecidas. O principal objetivo deste artigo é oferecer uma breve introdução ao paradoxo da conhecibilidade, apresentar brevemente algumas estratégias de solução para esse paradoxo, e discutir mais detalhadamente uma dessas tentativas de solução, viz. a tentativa de solução avançada por Jonathan Kvanvig (1996, 2006), de acordo com a qual a demonstração do paradoxo incorre em uma falácia modal.

Palavras-chave

Filosofia da Lógica. Paradoxo da Conhecibilidade. Falácias modais.

I - Agradeço a Juliana Faccio Lima e Sagid Salles pela leitura, revisões e pelos valiosos comentários que contribuíram para deixar o presente artigo bastante melhor.

Abstract

The subject matter of this paper is the knowability paradox. This paradox is a logical proof for the claim that if all truths are knowable, then all truths are known. The main goal of the paper is to offer a brief introduction to the knowability paradox, to present some strategies to solve it, and to discuss in more detail one of these attempts to solve the paradox, viz. Kvanvig's (1996, 2006) solution, according to which the proof in the knowability paradox relies on a modal fallacy.

Key-words

Philosophy of Logic. Knowability Paradox. Modal fallacies.

1. Introdução: o paradoxo, os problemas que ele levanta e um pouco de sua história

Paradoxos são argumentos aparentemente válidos, com premissas aparentemente verdadeiras, mas com uma conclusão aparentemente falsa. O que há de paradoxal em um paradoxo é que um argumento realmente válido, e com premissas realmente verdadeiras, não pode realmente possuir uma conclusão falsa.

Existem ao menos três maneiras de solucionar um paradoxo. A primeira é argumentar que, apesar do argumento ser aparentemente válido, ele não é realmente válido. Nesse caso, devemos apresentar bons motivos para rejeitar alguma regra de inferência utilizada no argumento, e explicar por que inicialmente essa regra de inferência nos parecia válida. A segunda é argumentar que, apesar de todas as premissas aparentemente serem verdadeiras, alguma delas não é realmente verdadeira. Nesse caso, devemos apresentar bons motivos para rejeitar alguma das premissas, e explicar porque inicialmente essa premissa nos parecia verdadeira. A terceira é argumentar que, apesar da conclusão aparentemente ser falsa, ela não é realmente falsa. Nesse caso, devemos apresentar bons motivos para aceitarmos a

conclusão, e explicar porque inicialmente essa conclusão nos parecia falsa.

O paradoxo da concecibilidade é uma demonstração lógica da afirmação de que *se todas as verdades são conhecíveis, então todas as verdades são conhecidas*. Antes de qualquer coisa, vale a pena clarificar um pouco essa afirmação.

Dizer que todas as verdades são conhecíveis é dizer que *se uma proposição é verdadeira, então é possível que essa proposição seja conhecida*. Supondo que a proposição «o número de batatas fritas que comi ao longo de minha vida é par» seja verdadeira, essa não seria uma verdade conhecida, pois aparentemente ninguém se importa com o número de batatas fritas que comi ao longo de minha vida. Mas seria uma verdade concebível, pois em princípio é possível que alguém, com bastante paciência e com interesses bastante estranhos, tivesse se dado ao trabalho de contar o número batatas fritas que comi ao longo de minha vida. Assim, dizer que todas as verdades são conhecíveis é dizer que isso ocorre com todas as proposições verdadeiras: independentemente de cada uma das proposições verdadeiras ser ou não ser efetivamente conhecida, cada uma delas pode em princípio ser conhecida. A essa afirmação chamaremos “tese da concecibilidade”, que pode ser formalizada assim²:

$$\forall p (p \rightarrow \Diamond Kp) \quad (\text{Tese da concecibilidade})$$

Dizer que *todas as verdades são conhecidas* é dizer que *se uma proposição é verdadeira, então essa proposição é efetivamente conhecida*. A essa afirmação chamaremos “tese da onisciência”, que pode ser formalizada assim:

$$\forall p (p \rightarrow Kp) \quad (\text{Tese da onisciência})$$

Nesses termos, o paradoxo da concecibilidade é uma suposta demonstração ou prova lógica de que se a tese da concecibilidade é verdadeira, então a tese da onisciência também é verdadeira. Esse resultado pode ser formalizado assim:

$$\forall p (p \rightarrow \Diamond Kp) \rightarrow \forall p (p \rightarrow Kp)$$

2 - Como é usual em lógica modal e epistêmica, o símbolo “ \Diamond ” deve ser lido como “é possível que” e o símbolo “K” deve ser lido como “é conhecido que”.

O principal objetivo deste artigo é oferecer uma breve introdução ao paradoxo da conhecibilidade. Nesta seção, começo por apresentar a demonstração desse paradoxo, os problemas que esse paradoxo levanta, e termino falando um pouco de sua história. Na segunda seção, apresento brevemente algumas estratégias de solução para esse paradoxo. O objetivo central da segunda seção é familiarizar o leitor com algumas estratégias gerais que uma tentativa de solução para o paradoxo pode seguir. Finalmente, na terceira seção, discuto mais detalhadamente uma dessas tentativas de solução, viz. a tentativa de solução avançada por Kvanvig (1996, 2006), de acordo com a qual a demonstração do paradoxo incorre em uma falácia modal.

1.1. A demonstração envolvida no paradoxo da conhecibilidade

Em termos informais, a linha de raciocínio que leva à conclusão do paradoxo da conhecibilidade é a seguinte: Suponha que todas as verdades são conhecíveis e suponha que existe uma verdade desconhecida. Pela segunda suposição, é verdade que uma proposição específica é verdadeira e desconhecida. Pela primeira suposição, é possível conhecer que essa proposição é verdadeira e desconhecida. Mas isso é contraditório, pois conhecer que uma proposição é verdadeira e desconhecida exige conhecer que essa proposição é verdadeira e conhecer que essa proposição é desconhecida. Mas se conhecemos que essa proposição é verdadeira, então essa proposição não é desconhecida. Logo, por redução ao absurdo, se todas as verdades são conhecíveis, então todas as verdades são conhecidas.

Em termos formais, a demonstração do paradoxo da conhecibilidade é uma demonstração por redução ao absurdo: Começamos supondo que a conclusão é falsa, i.e. que a condicional «*se todas as verdades são conhecíveis, então todas as verdades são conhecidas*» é falsa. Em seguida derivamos uma contradição dessa suposição, e isso nos permite inferir que, contrariamente à nossa suposição inicial, a condicional «*se todas as verdades são conhecíveis, então todas as verdades são conhecidas*» é, afinal, verdadeira. Como uma condicional é falsa se, e somente se, sua antecedente é verdadeira e sua consequente é falsa, então começamos supondo que todas as verdades são conhecíveis, apesar de algumas verdades serem desconhecidas:

$$\text{Sup} \quad (1) \quad \forall p (p \rightarrow \diamond Kp)$$

$$\text{Sup (2) } \exists p(p \wedge \neg Kp)$$

De acordo com nossa suposição (2), existe uma proposição específica que é verdadeira e desconhecida. Seja qual for essa proposição verdadeira e desconhecida, vamos chamá-la de p_1 :

$$(3) \quad p_1 \wedge \neg Kp_1$$

De acordo com (1), se uma proposição é verdadeira, então é possível conhecer que ela é verdadeira. Como a proposição « p_1 é verdadeira e desconhecida» é uma proposição como outra qualquer, então a suposição (1) nos permite inferir que se a proposição « p_1 é verdadeira e desconhecida» é verdadeira, então é possível conhecer que a proposição « p_1 é verdadeira e desconhecida» é verdadeira:

$$(4) \quad (p_1 \wedge \neg Kp_1) \rightarrow \diamond K(p_1 \wedge \neg Kp_1)$$

Esse passo pode parecer estranho, mas a princípio é apenas uma instanciação universal como outra qualquer. A ideia central é que a tese da conhecibilidade é uma tese sobre todas as proposições, viz. é a tese de que se uma proposição é verdadeira, então ela é conhecível. Como a tese da conhecibilidade é uma tese sobre todas as proposições, e como a proposição « $p_1 \wedge \neg Kp_1$ » é uma proposição como outra qualquer, então a tese da conhecibilidade nos permite inferir que se essa proposição é verdadeira, então ela é conhecível.

A partir das fórmulas (3) e (4), por *modus ponens*, podemos inferir que é possível conhecer que p_1 é verdadeira e desconhecida:

$$(5) \quad \diamond K(p_1 \wedge \neg Kp_1)$$

No entanto, podemos demonstrar que a fórmula (5) é logicamente falsa, utilizando apenas algumas regras da lógica clássica, modal e epistêmica. Para demonstrar que a fórmula (5) é logicamente falsa, começamos com a suposição de que a proposição « $p_1 \wedge \neg Kp_1$ » é conhecida:

$$\text{Sup (6) } K(p_1 \wedge \neg Kp_1)$$

Uma regra de inferência da lógica epistêmica é a distributividade do operador de conhecimento sobre conjunções. A ideia dessa regra é que se conhecemos uma conjunção, então também conhecemos cada lado da conjunção separadamente, e.g. se conhecemos que a neve é branca e a grama é verde, então conhecemos que a neve é branca e conhecemos que a grama é verde. Portanto, podemos inferir que é conhecido que a proposição p_1 é verdadeira e que é conhecido que não é conhecido que a proposição p_1 é verdadeira:

$$(7) \quad Kp_1 \wedge K\neg Kp_1$$

Outra regra de inferência da lógica epistêmica é a factividade do conhecimento. A ideia dessa regra é que se conhecemos que uma proposição é verdadeira, então essa proposição é verdadeira, e.g. se sabemos que a neve é branca, então a neve é branca. Assim, aplicando essa regra apenas ao lado direito da conjunção em (7), podemos inferir que é conhecido que a proposição p_1 é verdadeira e não é conhecido que a proposição p_1 é verdadeira:

$$(8) \quad Kp_1 \wedge \neg Kp_1$$

Como (8) é uma contradição, e como derivamos (8) a partir da suposição (6), por redução ao absurdo, podemos inferir que (6) é falsa:

$$(9) \quad \neg K(p_1 \wedge \neg Kp_1)$$

Como demonstramos que a fórmula (9) é verdadeira recorrendo apenas a regras de inferência, então a fórmula (9) é logicamente verdadeira. E como tudo que é logicamente verdadeiro é necessariamente verdadeiro, então podemos inferir que a fórmula (9) é necessariamente verdadeira:

$$(10) \quad \Box \neg K(p_1 \wedge \neg Kp_1)$$

Uma regra da lógica modal é a regra de interdefinição dos operadores modais. A ideia dessa regra é que “necessariamente não é o caso que ...” é equivalente a “não é possível que ...”, e.g. “necessariamente a neve não é preta” é equivalente a “não é possível que a neve seja preta”. A partir dessa regra e da fórmula (10), podemos inferir que não é possível conhecer que a proposição p_1 é verdadeira e desconhecida:

$$(11) \quad \neg \Diamond K(p_1 \wedge \neg Kp_1)$$

No entanto, (11) contradiz diretamente (5). Nesse caso, como começamos supondo que nossa conclusão é falsa, i.e. que a condicional «*se todas as verdades são conhecíveis, então todas as verdades são conhecidas*» é falsa, e como derivamos uma contradição dessa suposição, então, por redução ao absurdo, podemos inferir que a condicional «*se todas as verdades são conhecíveis, então todas as verdades são conhecidas*» deve ser verdadeira:

$$(12) \quad \forall p (p \rightarrow \Diamond Kp) \rightarrow \forall p (p \rightarrow Kp)$$

Que é exatamente a conclusão do paradoxo da conhecibilidade que estávamos querendo demonstrar.

1.2. Os problemas levantados pelo paradoxo da conhecibilidade

O paradoxo da conhecibilidade levanta ao menos dois problemas bastante diferentes. O primeiro problema é que essa demonstração estabelece que é logicamente verdadeiro que *se a tese da conhecibilidade é verdadeira, então a tese da onisciência também é verdadeira*, mas intuitivamente essa condicional não é uma verdade lógica. Em outras palavras, intuitivamente, a tese da conhecibilidade não implica logicamente a tese da onisciência.

É importante notar que a estranheza dessa conclusão não tem a ver com o valor de verdade efetivo da tese da conhecibilidade ou da tese da onisciência. Antes, a estranheza dessa

conclusão tem a ver com a relação lógica entre os valores de verdade dessas teses. Por um lado, mesmo que a tese da conhecibilidade e a tese da onisciência sejam ambas verdadeiras, continuaria sendo estranho que a primeira implicasse logicamente a segunda. Por outro lado, mesmo que a tese da conhecibilidade e a tese da onisciência sejam ambas falsas, também continuaria sendo estranho que a primeira implicasse logicamente a segunda.

Uma maneira de tornar mais clara a estranheza da conclusão é desenvolver um pouco mais uma de suas consequências. De acordo com a conclusão, a tese da conhecibilidade implica logicamente a tese da onisciência:

$$\forall p(p \rightarrow \diamond Kp) \vdash \forall p(p \rightarrow Kp)$$

Trivialmente, a tese da onisciência implica logicamente a tese da conhecibilidade, pois se todas as verdades são conhecidas, então obviamente todas as verdades são conhecíveis:

$$\forall p(p \rightarrow Kp) \vdash \forall p(p \rightarrow \diamond Kp)$$

A partir desses dois resultados, podemos concluir que a tese da conhecibilidade é logicamente equivalente à tese da onisciência:

$$\forall p(p \rightarrow \diamond Kp) \equiv \forall p(p \rightarrow Kp)$$

E aceitar isso significa aceitar que a tese da conhecibilidade possui as mesmas condições de verdade que a tese da onisciência. Mas isso é um resultado bastante estranho. A tese da conhecibilidade é uma afirmação modal, i.e. não é uma afirmação acerca de como as coisas efetivamente são, mas sim uma afirmação acerca de como as coisas poderiam ser. Mas a tese da onisciência não é uma afirmação acerca de como as coisas poderiam ser, mas sim uma afirmação acerca de como as coisas efetivamente são. Portanto, nosso último resultado estabelece que uma afirmação acerca de como as coisas poderiam ser possui as mesmas condições de verdade que uma afirmação acerca de como as coisas efetivamente são, e isso é bastante estranho.

O segundo problema levantado pelo paradoxo da conhecibilidade é um problema

específico para aqueles que defendem algumas teorias anti-realistas sobre a natureza da verdade. Grosso modo, uma teoria anti-realista sobre uma determinada área é uma teoria que sustenta que a verdade ou falsidade de afirmações nessa área depende de algum modo substancial de nossas crenças, práticas linguísticas, esquemas conceituais, ou algo do gênero. Por exemplo, um realista sobre verdades matemáticas tipicamente sustenta que números são entidades abstratas que existem independentemente de nossas crenças. Na medida em que os números são os objetos responsáveis pelas verdades matemáticas, então essas também seriam largamente independentes de nossas crenças, práticas linguísticas, esquemas conceituais e coisas do gênero. Contrariamente, um anti-realista tipicamente sustenta que números são construções humanas de algum tipo e que a ideia de verdade matemática deve ser reduzida à ideia de demonstração matemática, que também deve ser entendida como uma construção humana de algum tipo. Portanto, essas verdades seriam largamente dependentes de nossas crenças, práticas linguísticas, esquemas conceituais ou algo do gênero.

Mas como o paradoxo da conhecibilidade pode ser problemático para uma teoria anti-realista? Acontece que uma consequência usual de sustentar que verdades em uma determinada área são dependentes de nossas crenças, práticas linguísticas ou esquemas conceituais é que o conceito de verdade se torna um conceito largamente epistêmico. Por exemplo, a teoria verificacionista do significado sustenta que frases possuem sentido apenas quando podem, em princípio, ser verificadas ou falsificadas por evidências empíricas. Essa é uma teoria anti-realista típica: o conceito de significado é explicado em termos de verificabilidade em princípio e o conceito de verdade é explicado em termos de verificação efetiva, ou ideal, ou algo nessa linha. Assim, o conceito de verdade, que parecia um conceito não-epistêmico, é explicado em termos de verificação, que é um conceito paradigmaticamente epistêmico. Além disso, uma consequência usual de sustentar que o conceito de verdade é um conceito largamente epistêmico é que todas as verdades se tornam conhecíveis. Considere novamente o caso do verificacionismo: se uma frase possui sentido apenas se pode, em princípio, ser verificada, e como uma frase é verdadeira apenas se possui sentido, então uma frase é verdadeira apenas se pode, em princípio, ser verificada. Mas se uma frase pode, em princípio, ser verificada, então ela pode, em princípio, ser conhecida. Portanto, uma consequência do verificacionismo é que todas as frases que são verdadeiras podem, em princípio, ser conhecidas.

Agora podemos entender como o paradoxo da conhecibilidade pode ser problemático para uma teoria anti-realista. Como mencionado, uma das maneiras de solucionar um paradoxo é aceitar sua conclusão, i.e. aceitar que se todas as verdades são conhecíveis, então todas as verdades são conhecidas. Se adotarmos essa estratégia, então poderemos utilizar

essa conclusão em um argumento por redução ao absurdo contra as teorias anti-realistas que se comprometem com a ideia de que todas as verdades são conhecíveis. Em linhas gerais, o argumento é o seguinte: Se aceitarmos a conclusão do paradoxo da conhecibilidade, então todas as verdades são conhecíveis apenas se todas as verdades forem conhecidas. Nesse caso, as teorias anti-realistas que aceitam que todas as verdades são conhecíveis precisam aceitar também que todas as verdades são conhecidas. Como essa última consequência é obviamente falsa, então qualquer teoria que possua essa consequência deve ser rejeitada como igualmente falsa, e portanto qualquer teoria anti-realista que tenha como consequência que todas as verdades são conhecíveis deve ser rejeitada como falsa.

É importante distinguir claramente entre esses dois problemas. O primeiro problema é um problema mais geral, que tem a ver com nossa intuição pré-teórica de que a tese da conhecibilidade não implica logicamente a tese da onisciência, e de que essas duas teses possuem condições de verdade diferentes. O segundo problema é um problema mais específico, que tem a ver com algumas teorias anti-realistas que possuem como consequência a tese de que todas as verdades são conhecíveis. Assim, ao avaliar as soluções para o paradoxo da conhecibilidade, é importante manter claro se a solução em questão é uma solução para o primeiro ou para o segundo problema. Pois uma solução para o primeiro problema não é automaticamente uma solução para o segundo problema, e.g. uma solução para o primeiro problema pode depender de suposições incompatíveis com as teorias anti-realistas. E uma solução para o segundo problema não é automaticamente uma solução para o primeiro problema, e.g. uma reformulação do anti-realismo que não se comprometa com a tese da conhecibilidade pode não servir para solucionar o primeiro problema.

1.3. Um pouco da história do paradoxo da conhecibilidade

Até agora apresentei a demonstração envolvida no paradoxo da conhecibilidade como sendo um paradoxo genuíno, além de ser também um problema em potencial para algumas teorias anti-realistas. Mas nem todos os autores encaram essa demonstração como um paradoxo genuíno, isto é, como um paradoxo a ser resolvido e não apenas como uma demonstração de um resultado a ser aceito. O primeiro autor a apresentar a demonstração foi Frederic Fitch (1963). Nesse artigo, Fitch apresenta algumas demonstrações de alguns teoremas, dentre os quais está o que ele chama de “Teorema 5”. Esse teorema afirma que “se existe alguma proposição verdadeira que ninguém sabe (ou soube ou saberá) ser verdadeira, então existe uma proposição que ninguém pode saber ser verdadeira”. A contrapositiva desse

teorema, assim como sua demonstração, é o que veio a ser conhecido como o paradoxo da conhecibilidade. No entanto, Fitch não percebe coisa alguma de contra intuitivo nesse resultado, e apresenta esse teorema, assim como sua demonstração, como um resultado a ser aceito, e não como um paradoxo a ser resolvido.

O artigo de Fitch não foi um dos mais populares por algum tempo, pois a demonstração de Fitch não foi mencionada sequer uma vez por pelo menos dez anos desde sua publicação. Os primeiros autores a redescobrirem e a perceberem algumas consequências filosóficas mais profundas da demonstração de Fitch foram W. D. Hart e Colin McGinn (Hart e McGuinn, 1976 e Hart, 1979). Eles se referem a essa demonstração como uma “pérola lógica negligenciada”, capaz de refutar definitivamente o verificacionismo. O argumento apresentado por eles contra o verificacionismo é essencialmente o mesmo que apresentei na seção anterior. Apesar de Hart e McGinn perceberem as consequências dessa demonstração para o verificacionismo, nenhum deles parece ter percebido claramente que sua conclusão era pré-teoricamente contra-intuitiva, e, portanto, que deveria ser considerada primeiramente um paradoxo a ser resolvido, e não simplesmente um resultado a ser aceito.

O primeiro autor a discutir se a demonstração de Fitch é um paradoxo genuíno, e a discutir em mais pormenores suas consequências para o verificacionismo, foi John Mackie (1980). Apesar de Mackie conceder que a conclusão dessa demonstração inicialmente parece surpreendente, ele argumenta que ela não é realmente falsa, e procura explicar nossa surpresa inicial. A ideia de Mackie é que não deve ser surpreendente que operadores factivos possam ser utilizados para criar exemplos auto-refutantes. Considere este caso: Pode ser verdade que não estou escrevendo, e posso escrever que não estou escrevendo, mas não posso escrever *verdadeiramente* que não estou escrevendo. Acontece que o ato de escrever a frase “não estou escrevendo” falsifica a frase que está sendo escrita. Apesar de isso não ser um impedimento para escrever essa frase, isso é um impedimento para escrever *verdadeiramente* a frase. Agora considere o caso do paradoxo da conhecibilidade: Pode ser verdade que «a neve é branca e ninguém conhece que a neve é branca», mas não é possível conhecer que «a neve é branca e ninguém conhece que a neve é branca». Acontece que o ato de conhecer a proposição «a neve é branca e ninguém conhece que a neve é branca» falsifica a proposição, e isso é um impedimento para conhecer a proposição. Mackie defende que esses dois casos são essencialmente similares, e como o primeiro caso não deve ser surpreendente, então o segundo caso também não deve ser surpreendente.

Penso que essa tentativa de dissolver a surpresa da conclusão de nosso paradoxo não funciona. A explicação de Mackie não dissolve a surpresa de que a tese segundo a qual todas as verdades podem ser conhecidas é logicamente equivalente à tese segundo a

qual todas as verdades efetivamente são conhecidas. A explicação de Mackie pode tornar intuitiva a demonstração dessa conclusão: Se todas as verdades pudessem ser conhecidas e se existisse uma verdade desconhecida, então seria possível conhecer a conjunção de que essa verdade é uma verdade e que essa verdade é desconhecida. Mas isso não é possível, pois o processo de conhecer essa conjunção falsifica a conjunção. No entanto, que a explicação de Mackie torne intuitiva essa demonstração não ajuda em nada a solucionar o paradoxo, pois a demonstração é intuitiva desde o começo, e isso é parte do problema: Temos uma demonstração aparentemente válida, na qual todas as inferências aparentemente são válidas, e uma conclusão surpreendente. Apresentar uma explicação que reitera a intuitividade da demonstração não ajuda em nada a dissolver a surpresa da conclusão. Se queremos dissolver a surpresa dessa conclusão, então devemos identificar os motivos pelos quais a conclusão parece surpreendente, e apresentar argumentos independentes para mostrar que esses motivos são equivocados.

A primeira autora a tratar a demonstração do paradoxo da conhecibilidade como um paradoxo genuíno foi Dorothy Edgington (1985), mas por motivos diferentes daqueles que apresentei. De acordo com a maneira que encaro essa demonstração, ela é paradoxal em virtude de estabelecer uma equivalência lógica entre a tese da conhecibilidade e a tese da onisciência. Nesse caso, a surpresa da conclusão não tem a ver com teoria alguma sobre a natureza da verdade. De acordo com a maneira que Edgington encara essa demonstração, ela é paradoxal em virtude de reduzir ao absurdo de modo muito simples e muito direto diversas teorias sofisticadas sobre a natureza da verdade. Nas palavras de Edgington, “mesmo um realista completo deve suspeitar que a longa tradição filosófica à qual ele se opõe não possa ser tão rapidamente reduzida ao absurdo”³. Nesse caso, a surpresa da conclusão tem a ver diretamente com essas teorias anti-realistas sobre a natureza da verdade.

Penso que essa não é a melhor maneira de encarar essa demonstração. Nas palavras de Williamson, “se os proponentes dessas teorias não perceberam um contra-exemplo simples, isso é um embaraço, não um paradoxo”⁴. Mesmo porque, se isso fosse motivo suficiente para considerarmos a demonstração paradoxal, então qualquer argumento simples e direto contra qualquer teoria que defendemos poderia ser encarado como um paradoxo a ser resolvido, e não como uma refutação de uma teoria a ser abandonada.

O primeiro autor a tratar essa demonstração como um paradoxo genuíno pelos motivos que apresentei foi Jonathan Kvanvig (1995, 2006). Kvanvig argumenta extensamente que a demonstração de Fitch é paradoxal em primeiro lugar por estabelecer, contrariamente a nossas intuições, uma equivalência entre a tese da conhecibilidade e a tese da onisciência.

3 - Edgington (1985), p. 568, tradução minha.

4 - Williamson (2000), p. 271, tradução minha.

Em virtude disso, e não em virtude de reduzir ao absurdo algumas teorias anti-realistas, ela deve ser encarada primeiramente como um paradoxo a ser resolvido, e não como um resultado a ser aceito.

Agora que estamos mais familiarizados com a demonstração do paradoxo da conhecibilidade, com os problemas que esse paradoxo levanta, e com um pouco de sua história, podemos começar a considerar algumas das soluções propostas.

2. Estratégias de solução: rejeitar a validade ou aceitar a conclusão

Como mencionei, existem ao menos três maneiras de solucionar um paradoxo. A primeira é argumentar que o argumento não é realmente válido; a segunda é argumentar que alguma das premissas do argumento não é realmente verdadeira; e a terceira é aceitar que a conclusão do argumento é realmente verdadeira. No entanto, na medida em que o argumento envolvido no paradoxo da conhecibilidade é uma demonstração lógica, i.e. é um argumento sem premissas, que depende apenas de inferências logicamente válidas, então temos realmente apenas duas opções viáveis, viz. ou rejeitamos a validade de alguma inferência ou aceitamos a conclusão do argumento.

Seguindo a linha de rejeitar a validade do argumento, existem ao menos duas estratégias gerais que podemos adotar. A primeira é aceitar que todas as regras de inferência utilizadas no argumento são válidas, mas argumentar que alguma regra está sendo aplicada incorretamente. A segunda é argumentar que alguma regra de inferência utilizada no argumento é inválida.

Nesta seção descrevo brevemente algumas tentativas de solução que exemplificam essas estratégias. O objetivo desta seção não é fornecer explicações e discussões detalhadas de cada uma das soluções, mas apenas familiarizar o leitor com as direções gerais que as diferentes soluções para o paradoxo podem tomar, e apontar ao leitor as referências bibliográficas relevantes.

2.1. A solução de Kvanvig: falácias modais

Kvanvig (1996, 2006) adota a linha de argumentar que a demonstração do paradoxo da

conhecibilidade é inválida. Mas não porque alguma regra de inferência utilizada é inválida, mas sim porque uma dessas regras está sendo aplicada incorretamente.

De acordo com Kvanvig, a demonstração do paradoxo da conhecibilidade envolve uma falácia modal. No passo (4) da demonstração, instanciamos a conjunção “ $p_1 \wedge \neg Kp_1$ ” como o valor da variável “ p ” na tese da conhecibilidade, i.e. na fórmula “ $\forall p(p \rightarrow \Diamond Kp)$ ”. Apesar de Kvanvig aceitar que instanciações universais são geralmente válidas, ele nota que precisamos ser cuidadosos ao aplicarmos essa regra de inferência quando estamos lidando com fórmulas que contenham operadores modais. Acontece que substituições de expressões modalmente flexíveis no âmbito de operadores modais são geralmente inválidas. Desse ponto de partida, Kvanvig argumenta que a conjunção “ $p_1 \wedge \neg Kp_1$ ” é uma expressão modalmente flexível, e como ela é substituída no âmbito de um operador modal na demonstração do paradoxo, temos motivos para pensar que esse passo envolve uma aplicação ilegítima da regra de instanciação universal.

Como discutirei essa solução em mais detalhes na próxima seção, deixo essa abordagem de lado por hora.

2.2. A solução de Linsky: a teoria dos tipos

Linsky (2009) segue a mesma linha de defender que a demonstração do paradoxo da conhecibilidade é inválida argumentando que alguma regra de inferência utilizada na demonstração está sendo aplicada incorretamente⁵. Mais especificamente, Linsky argumenta que o problema com a demonstração de Fitch está na passagem do passo (8) para o passo (9). No passo (8), assumimos que a fórmula “ $Kp_1 \wedge \neg Kp_1$ ” é contraditória, e usamos isso para inferir, por redução ao absurdo, que a fórmula “ $K(p_1 \wedge \neg Kp_1)$ ” não pode ser verdadeira. Linsky, apesar de aceitar que argumentos por redução ao absurdo são geralmente válidos, defende que a fórmula (8) não é realmente contraditória.

A ideia central aqui é que precisamos distinguir claramente entre os conceitos de *conhecimento de primeira ordem*, i.e. conhecimento de verdades ordinárias que não contenham ocorrência alguma do conceito de conhecimento, *conhecimento de segunda ordem*, i.e. conhecimento sobre aquilo que é conhecido, *conhecimento de terceira ordem*, i.e. conhecimento sobre conhecimento sobre aquilo que é conhecido, e assim por diante. Quando dizemos que uma proposição específica p_1 é verdadeira e desconhecida, estamos

⁵ - Church (2009) foi o primeiro a sugerir a linha de solução que Linsky defende, mas aqui discuto Linsky pois ele foi o primeiro a desenvolver essa sugestão em mais pormenor.

lidando com conhecimento de um determinado nível, dependendo do conteúdo da proposição p_1 . Assuma que p_1 é uma verdade ordinária e portanto que estamos lidando com um conhecimento de primeira ordem. Usando o símbolo “ $K_{(1)}$ ” para representar esse conceito de conhecimento, devemos representar o caso assim:

$$p_1 \wedge \neg K_{(1)}p_1$$

Mas quando dizemos que é conhecido que essa proposição específica é verdadeira e desconhecida, estamos lidando com um conhecimento sobre aquilo que é conhecido, i.e. com um conhecimento de segunda ordem. Usando o símbolo “ $K_{(2)}$ ” para representar o conceito de conhecimento de segunda ordem, devemos representar o caso assim:

$$K_{(2)}(p_1 \wedge \neg K_{(1)}p_1)$$

Dessa fórmula, como na demonstração original, ainda podemos derivar a seguinte consequência:

$$K_{(2)}p_1 \wedge \neg K_{(1)}p_1$$

Mas agora essa fórmula não é mais contraditória, pois estamos lidando com conhecimentos de diferentes ordens. Portanto, a demonstração do paradoxo da conhecibilidade é bloqueada e o paradoxo estaria resolvido.

Paseau (2008, 2009), Halbach (2008) e Carrara & Fassio (2011) também discutem essa linha de solução.

2.3. A solução de Beall: o dialeteísmo

J. C. Beall (2000) adota uma linha mais radical e argumenta que a demonstração do paradoxo da conhecibilidade é inválida porque uma das regras de inferência utilizadas é inválida⁶.

6 - Routley (1981) foi o primeiro a sugerir essa solução para o paradoxo da conhecibilidade. Aqui discuto Beall (2000) pois

J. C. Beall e G. Priest são famosos por defender uma posição conhecida como “dialeteísmo”, i.e. a tese de que algumas contradições são verdadeiras. A motivação usual para essa posição é a existência de alguns paradoxos particularmente difíceis de serem resolvidos. O exemplo paradigmático aqui é o paradoxo do mentiroso: Considere a frase “Esta frase é falsa”. Se ela é verdadeira, então o que ela diz é o caso. Como ela diz de si própria que ela é falsa, então ela é falsa. Mas se ela é falsa, então o que ela diz não é o caso. Como ela diz de si própria que ela é falsa, então ela é verdadeira. Portanto, a frase é verdadeira se e somente se for falsa. Contradição.

O paradoxo do mentiroso é um dos paradoxos mais antigos dos quais temos conhecimento, e mesmo depois de mais de 2.000 anos, ainda não temos uma maneira completamente satisfatória de bloquear sua demonstração. Mas como sabemos, encontrar algo de errado com o argumento é apenas uma das maneiras de solucionar um paradoxo. Ainda temos a opção de aceitar o argumento e sua conclusão. A sugestão dos defensores do dialeteísmo é que, depois de tanto tempo procurando e não encontrando algo de errado com o argumento envolvido no paradoxo do mentiroso, talvez seja a hora de começar a considerar a possibilidade de que o argumento seja sólido e aceitar sua conclusão: a frase “Esta frase é falsa” é verdadeira e falsa ao mesmo tempo.

Se algumas contradições realmente podem ser verdadeiras, então devemos ser mais cuidadosos com argumentos por redução ao absurdo, como é o caso do argumento do paradoxo da conhecibilidade. Afinal, talvez a contradição “ $Kp_1 \wedge \neg Kp_1$ ”, que usamos para completar uma redução ao absurdo em nosso paradoxo, seja uma dessas contradições verdadeiras. Mas existe de fato algum motivo independente para pensarmos que essa contradição em particular seja uma dessas contradições verdadeiras?

Beall (2000) argumenta que sim. Do mesmo modo que o paradoxo do mentiroso nos dá razões para pensar que alguns conceitos semânticos geram algumas contradições verdadeiras, o paradoxo do conhecedor (não confundir com o paradoxo da conhecibilidade) nos dá razões para pensar que alguns conceitos epistêmicos também geram algumas contradições verdadeiras. Grosso modo, o paradoxo do conhecedor é o seguinte: Considere a frase “Ninguém sabe que esta frase é verdadeira”. Assuma que sabemos que essa frase é verdadeira. Como o conceito de conhecimento é factivo, então ela é verdadeira. Assuma que não sabemos que ela é verdadeira. Como ela afirma justamente que não sabemos que ela é verdadeira, então ela é verdadeira. Portanto, em qualquer caso, a frase “Ninguém sabe que esta frase é verdadeira” é verdadeira. Mas isso é uma demonstração de que a frase é verdadeira. Como acabamos de demonstrar que ela é verdadeira, então sabemos que ela é verdadeira. Mas se sabemos que ela é verdadeira, então o que ela diz é ele foi o primeiro a desenvolver essa sugestão em mais pormenor.

o caso, e ela diz que não sabemos que ela é verdadeira. Portanto, sabemos e não sabemos que ela é verdadeira. Contradição. Note que a contradição que obtemos como conclusão no paradoxo do conhecedor é bastante similar com a tipo de contradição que usamos na redução do paradoxo da conhecibilidade, viz. é uma contradição da forma “ $Kp_1 \wedge \neg Kp_1$ ”. Portanto, se tivermos motivos fortes o suficiente para pensar que o dialeteísmo é também uma boa solução para o paradoxo do conhecedor, parece que temos alguma motivação independente para pensar que a contradição “ $Kp_1 \wedge \neg Kp_1$ ” na demonstração do paradoxo da conhecibilidade pode ser um exemplo de uma contradição verdadeira.

Beall (2009) e Priest (2009) também discutem e desenvolvem essa linha de solução em mais detalhes.

2.4. A solução de Williamson e Dummett: o intuicionismo

Ao considerar diferentes soluções para o paradoxo da conhecibilidade, precisamos ser cuidadosos em distinguir claramente os dois problemas levantados pelo paradoxo. Como mencionei anteriormente, podemos encarar o paradoxo da conhecibilidade como um problema para algumas versões do anti-realismo ou podemos encarar esse paradoxo como um paradoxo em si mesmo. Muitas vezes a solução para um desses problemas é independente da solução para o outro problema. Williamson (1982) e Dummett (2009), por exemplo, argumentam que se aceitamos uma lógica intuicionista, a conclusão do paradoxo da conhecibilidade deixa de ser contra-intuitiva. Além disso, eles também argumentam que os anti-realistas possuem motivações independentes para rejeitar a lógica clássica em favor de uma lógica intuicionista. Por isso, eles concluem que o paradoxo da conhecibilidade não deve ser particularmente problemático para os anti-realistas. Essa solução pode ser promissora como uma solução para o paradoxo da conhecibilidade enquanto um problema para os anti-realistas. Mas essa solução certamente não parece muito promissora como uma solução para o paradoxo da conhecibilidade enquanto um paradoxo em si mesmo. Pois ainda que os anti-realistas realmente tenham motivações independentes para aceitar uma lógica intuicionista em virtude de seu comprometimento com o anti-realismo, é bastante duvidoso que realistas tenham qualquer motivação independente para aceitar uma lógica intuicionista.

Williamson (1988, 1992), Percival (1990, 1991), DeVidi & Solomon (2001), Brogaard & Salerno (2002), Burmüdez (2009), Dummett (2009), Rasmussen (2009) Maffezoli, Naibo & Negri (2012) também discutem essa abordagem.

2.5. A solução de Tennant: proposições cartesianas

Tennant (1997) também oferece uma solução que, ainda que seja promissora para resolver o paradoxo da conhecibilidade enquanto um problema para os anti-realistas, não parece muito promissora para resolver o paradoxo da conhecibilidade enquanto um paradoxo em si mesmo. De acordo com Tennant, os anti-realistas não precisam se comprometer com a tese de que absolutamente todas as verdades são conhecíveis. Antes, os anti-realistas podem se comprometer apenas com uma versão devidamente restrita desse princípio. Sua sugestão é que os anti-realistas se comprometam apenas com a tese de que todas as verdades cartesianas são conhecíveis, onde uma verdade cartesiana é, grosso modo, uma verdade que pode consistentemente ser conhecida.

Ainda que essa solução seja viável para resolver o paradoxo da conhecibilidade enquanto um problema para os anti-realistas, ela certamente não parece muito promissora enquanto uma solução para o paradoxo da conhecibilidade enquanto um paradoxo. Pois ainda que a demonstração do paradoxo seja bloqueada quando consideramos a tese da conhecibilidade restrita para verdades cartesianas, a demonstração ainda será válida para a tese da conhecibilidade irrestrita, e isso continua sendo paradoxal.

Discussões, objeções e desenvolvimentos posteriores dessa abordagem incluem Tennant (2001a, 2001b, 2002, 2009, 2010), Hand & Kvanvig (1999), Hand (2003), Williamson (2000b, 2009), DeVidi & Kenyon (2003), Brogaard and Salerno (2002, 2006, 2008), e Rosenkranz (2004).

2.6. As soluções de Edgington, Fara e Fuhrmann: conhecimento possível de verdades atuais, capacidade para conhecer e conhecimento potencial

Como vimos, muitas vezes uma solução para o paradoxo da conhecibilidade enquanto um problema para algumas versões do anti-realismo é em larga medida independente de uma solução para esse paradoxo como um paradoxo em si mesmo. Ainda assim, muitas vezes a solução para o primeiro problema é de algum modo relevante para a solução do segundo problema. Edgington (1985), por exemplo, também argumenta que os anti-realistas não precisam se comprometer com a tese de que absolutamente todas as verdades são

conhecíveis. A sugestão de Edgington é que os anti-realistas se comprometam apenas com a tese de que, para cada proposição que é verdadeira no mundo atual, é possível conhecer que essa proposição é verdadeira no mundo atual⁷. Nessa mesma linha, Fara (2010) e Fuhmann (2014) argumentam que os anti-realistas não devem entender conhecibilidade em termos de conhecimento possível, mas sim em termos de capacidade para conhecer ou em termos de conhecimento potencial. De acordo com eles, suas respectivas reformulações evitam a demonstração do paradoxo, e portanto os anti-realistas podem evitar o paradoxo da conhecibilidade.

Uma vantagem dessas soluções é que, apesar delas serem construídas como soluções para o paradoxo da conhecibilidade enquanto um problema para os anti-realistas, essas mesmas soluções podem ser facilmente adaptadas para funcionar como uma solução para o paradoxo da conhecibilidade enquanto um paradoxo em si mesmo. Quando adaptadas para funcionar como uma solução para o paradoxo da conhecibilidade enquanto um paradoxo, essa linha de solução consiste em argumentar que, apesar de a conclusão ser surpreendente, ela é correta. A ideia geral aqui é a seguinte: Temos motivos independentes para pensar que as condições de verdade da tese da conhecibilidade, i.e. da tese segundo a qual todas as verdades são conhecíveis, não são corretamente capturadas pela fórmula “ $\forall p(p \rightarrow \Diamond Kp)$ ”. Uma dessas motivações é justamente que, quando representamos a tese da conhecibilidade através dessa fórmula, acabamos por gerar um paradoxo. Uma analogia pode ser esclarecedora aqui. Considere os paradoxos da condicional material. Quando representamos condicionais da linguagem natural da forma “se ϕ , então ψ ” através da condicional material “ $\phi \rightarrow \psi$ ”, acabamos por gerar certas consequências contra-intuitivas. Por exemplo, acabamos por ter como consequência que condicionais como “se $2 + 2 = 5$, então existe vida em outros planetas” e “se existe vida em outros planetas, então $2 + 2 = 4$ ” são verdadeiras. Mas essas condicionais certamente não parecem verdadeiras. Uma estratégia aqui é argumentar que as condições de verdade das condicionais da linguagem natural não são corretamente capturadas pela condicional material da lógica clássica. Portanto, precisamos encontrar uma maneira alternativa de representar as condições de verdade das condicionais da linguagem natural. Similarmente, quando representamos a tese da conhecibilidade através da fórmula “ $\forall p(p \rightarrow \Diamond Kp)$ ”, acabamos por gerar certas consequências contra-intuitivas. Por exemplo, acabamos por ter como consequência que a tese da conhecibilidade é equivalente à tese da onisciência. Mas essas teses certamente não parecem equivalentes. Uma estratégia análoga aqui é argumentar que as condições de verdade da tese da conhecibilidade não são corretamente capturadas pela fórmula “ $\forall p(p \rightarrow \Diamond Kp)$ ”. Portanto, precisamos encontrar uma maneira alternativa de representar as condições

7 - Schlesinger (1985) foi o primeiro a sugerir uma solução nessas linhas. Aqui discuto Edgington (1985) pois ela apresenta essa proposta mais claramente e em mais detalhes.

de verdade da tese da concecibilidade.

Outra motivação para pensar que as condições de verdade da tese da concecibilidade não são corretamente capturadas pela fórmula “ $\forall p(p \rightarrow \Diamond Kp)$ ” é a seguinte. Por um lado, o conceito de concecibilidade parece ser factivo, i.e. se uma proposição é concecível, então ela é verdadeira. Nossas intuições certamente apontam para essa direção: é estranho dizer que é concecível que a neve é preta, justamente porque é falso que a neve seja preta. Por outro lado, na tese da concecibilidade estamos representando o conceito de concecibilidade em termos de conhecimento possível. Isto é, estamos assumindo que expressões da forma “é concecível que ϕ ” devem ser entendidas como sinônimas de expressões da forma “é possível conhecer que ϕ ”. No entanto, expressões da forma “é possível conhecer que ϕ ” não são factivas, na medida em que não há nada de errado com a ideia de que existem possibilidades nas quais uma proposição, que atualmente é falsa, seja verdadeira e conhecida. Nesse caso, temos motivos para desconfiar da formulação da tese da concecibilidade em termos de conhecimento possível.

Esses argumentos dependem apenas das nossas intuições sobre a tese da concecibilidade e sobre o conceito de concecibilidade, e podem ser aceitos tanto por anti-realistas quanto por realistas. Portanto, se essa linha de argumento funciona como motivação para os anti-realistas reformularem a tese da concecibilidade, então parece que eles também deveriam motivar os realistas a reformularem a tese da concecibilidade. Nesse caso, se as soluções de Edgington, Fara e Fuhrmann são boas soluções para o paradoxo da concecibilidade enquanto um problema para os anti-realistas, então parece que elas também podem ser boas soluções para o paradoxo da concecibilidade enquanto um paradoxo.

Rabinowicz & Segerberg (1994) e Lindström (1997) desenvolvem alguns detalhes mais técnicos dessa abordagem. Williamson (1987a, 1987b, 2000, no prelo), Percival (1991), Edgington (2010), Fara (2010) e Fuhrmann (2014) também discutem essa linha de solução.

Agora que estamos mais familiarizados com algumas das tentativas de solução mais salientes na bibliografia sobre o paradoxo da concecibilidade, podemos discutir uma dessas soluções em mais detalhes.

2. A solução de Kvanvig: falácias modais

Como mencionado, a solução de Kvanvig consiste em argumentar que, apesar da

demonstração do paradoxo da concecibilidade ser aparentemente válida, ela não é realmente válida. O argumento de Kvanvig para rejeitarmos uma das inferências da demonstração pode ser formulado assim:

A.1. Todas as inferências que dependem da substituição de expressões modalmente flexíveis no âmbito de operadores modais são inferências inválidas.

A.2. Uma das inferências da demonstração do paradoxo da concecibilidade depende da substituição de uma expressão modalmente flexível no âmbito de um operador modal.

A.3. Logo, uma das inferências da demonstração do paradoxo da concecibilidade é uma inferência inválida.

Antes de discutirmos esse argumento em mais detalhes, apresentarei alguns conceitos que precisamos para entendê-lo melhor, viz. os conceitos de mundo possível, expressões referenciais modalmente rígidas e modalmente flexíveis, frases modalmente rígidas e modalmente flexíveis e falácias modais.

2.1. Expressões modalmente flexíveis e falácias modais

Começemos pelo conceito de mundo possível. O conceito de *mundo possível* serve para representar nossas afirmações sobre possibilidades e necessidades. Intuitivamente, um mundo possível é um modo como as coisas poderiam ser. Por exemplo, apesar de Sócrates ter sido grego, ele poderia ter sido egípcio, e nesse sentido podemos dizer que existe um mundo possível com respeito ao qual Sócrates foi egípcio.

Utilizando o conceito de mundo possível, nossas afirmações sobre *necessidades* podem ser representadas como quantificações universais sobre mundos possíveis. Assim, a afirmação de que é necessário que Sócrates seja mortal pode ser representada pela afirmação de que Sócrates é mortal com respeito a todos os mundos possíveis. Na mesma linha, nossas afirmações sobre *possibilidades* podem ser representadas como quantificações existenciais sobre mundos possíveis. Assim, a afirmação de que Sócrates poderia ter sido egípcio pode ser representada pela afirmação de que Sócrates é egípcio com respeito a *algum* mundo possível.

Os conceitos de *expressão modalmente rígida* e *expressão modalmente flexível* servem para caracterizar o comportamento de algumas expressões em frases com operadores modais. Uma expressão modalmente rígida é uma expressão que designa o mesmo objeto com respeito a qualquer mundo possível. Por exemplo, a expressão “Sócrates” aparentemente é uma expressão modalmente rígida. Quando utilizamos essa expressão para falar de um objeto com respeito a algum mundo possível, estamos sempre falando do mesmo objeto, viz. do objeto que foi um filósofo ateniense no mundo atual. Contrariamente, uma expressão modalmente flexível é uma expressão que designa diferentes objetos com respeito a diferentes mundos possíveis. Por exemplo, a expressão “o número de planetas” aparentemente é uma expressão modalmente flexível. Quando utilizamos essa expressão para falar de um objeto com respeito a um mundo possível no qual existem nove planetas, estamos falando do número nove, mas quando utilizamos essa expressão para falar de um objeto com respeito a um mundo possível no qual existem oito planetas, estamos falando do número oito.

Expressões modalmente flexíveis, quando utilizadas em alguns argumentos modais, produzem certos tipos de inferências problemáticas, usualmente chamadas de *falácias modais*. Um exemplo familiar desse tipo de inferência é o seguinte:

- (a.i) Para qualquer número n , se n é par, então é necessário que n seja par;
- (a.ii) Logo, se o *número de planetas* é par, então é necessário que o *número de planetas* seja par.

Esse tipo de inferência é problemática porque sua premissa parece verdadeira, a inferência parece válida, mas sua conclusão parece falsa. Mas claro que uma inferência válida com premissas verdadeiras não pode ter uma conclusão falsa. Em linhas gerais, o diagnóstico usual para nossas intuições conflitantes sobre essa inferência é que sua conclusão é ambígua entre uma leitura *de dicto* e uma leitura *de re*. Por um lado, se interpretarmos a conclusão segundo a leitura *de dicto*, então a conclusão é falsa, mas a inferência é inválida, apesar da aparência inicial de validade. Por outro lado, se interpretarmos a conclusão segundo a leitura *de re*, então a inferência é válida, mas a conclusão é verdadeira, apesar da aparência inicial de falsidade.

O quê exatamente são as leituras *de dicto* e *de re*? A ideia geral é que uma frase como “necessariamente o número de planetas é maior que cinco” pode ser interpretada como uma afirmação sobre um conceito, viz. que o conceito «o número de planetas» necessariamente

denota um objeto que possui a propriedade de ser maior que cinco, ou como uma afirmação diretamente sobre o objeto que esse conceito efetivamente denota, viz. que o número oito, que é o objeto que o conceito «o número de planetas» efetivamente denota, necessariamente possui a propriedade de ser maior que cinco. A primeira interpretação é a leitura *de re*, na qual uma propriedade está sendo atribuída diretamente a um objeto. A segunda interpretação é a leitura *de dicto*, na qual fazemos uma afirmação sobre o conceito.

Com essa distinção, podemos explicar o que está acontecendo com nosso argumento sobre o número de planetas. De acordo com a leitura *de dicto*, a conclusão da inferência afirma que para qualquer mundo possível w , o número de planetas com respeito à w possui a propriedade de ser par com respeito à w , se possuir a propriedade de ser par com respeito ao mundo atual. A ideia central é que estamos atribuindo uma propriedade ao conceito «o número de planetas», viz. a propriedade de necessariamente denotar um número par, se atualmente denotar um número par. Nesse caso, intuitivamente essa afirmação é falsa, pois esse conceito denota o número um com respeito a um mundo possível no qual todos os planetas com exceção de Saturno foram destruídos, apesar de denotar o número oito com respeito ao mundo atual. Mas nesse caso, a inferência é inválida, pois da afirmação de que qualquer número necessariamente possui a propriedade de ser par, se atualmente for par, não se segue a afirmação de que o conceito «o número de planetas» necessariamente denota um número par, se atualmente denotar um número par.

De acordo com a leitura *de re*, a conclusão da inferência afirma que para qualquer mundo possível w , o número de planetas com respeito ao mundo atual é par com respeito a esse mundo possível w , se for par com respeito ao mundo atual. Aqui estamos atribuindo uma propriedade diretamente ao objeto denotado pelo conceito «o número de planetas», viz. a propriedade de necessariamente ser par, se atualmente for par. Nesse caso, a inferência é válida, pois da afirmação de que qualquer número necessariamente é par, se atualmente for par, se segue que o objeto atualmente denotado pelo conceito «o número de planetas» é necessariamente par, se atualmente for par. Isso porque estamos atribuindo a uma propriedade diretamente ao objeto, que é o mesmo em todos os mundos possíveis. Mas nesse caso a conclusão da inferência é verdadeira, pois o objeto atualmente denotado pelo conceito «o número de planetas», viz. o número oito, possui a propriedade de necessariamente ser par, se atualmente for par.

Esse caso ilustra como a substituição de expressões modalmente flexíveis no âmbito de operadores modais resulta em inferências problemáticas, que produzem um tipo característico de invalidade, como é o caso da leitura *de dicto* de nosso exemplo.

2.2. Frases modalmente flexíveis e falácias modais

O último conceito que precisamos é o conceito de frase *modalmente flexível*. Essa ideia exige alguns esclarecimentos preliminares. Usualmente o tipo de expressões que caracterizamos como modalmente rígidas ou flexíveis são *expressões referenciais*, i.e. expressões que possuem a função de denotar objetos. Por exemplo, as expressões “Sócrates” e “o número de planetas” aparentam ser expressões referenciais, na medida em que possuem a função de denotar objetos. No entanto, a estratégia de Kvanvig é argumentar que a expressão “ $p_1 \wedge \neg Kp_1$ ” é modalmente flexível. Acontece que a expressão “ $p_1 \wedge \neg Kp_1$ ” não é uma expressão referencial, mas sim uma *frase*, i.e. não é uma expressão que possui a função de denotar objetos, mas sim de expressar proposições. Portanto, para entendermos o argumento de Kvanvig, precisamos esclarecer em que sentido frases, mesmo não sendo expressões referenciais, ainda podem ser modalmente rígidas ou flexíveis.

Apesar de não ser claro que podemos falar de frases designando proposições, e de ser menos claro ainda que podemos falar de frases designando flexivelmente proposições, penso que podemos definir um conceito mais geral que chamarei de “designação*” e que é suficiente para formular a ideia de que frases podem ser expressões modalmente rígidas ou modalmente flexíveis. Para o caso das expressões referenciais, a definição de designação* é a seguinte:

Uma expressão referencial “a” designa* um objeto x em uma frase como “Fa” se, e somente se, x é o objeto que deve satisfazer o predicado “F” para tornar a frase “Fa” verdadeira.

A ideia geral é que uma expressão referencial como “Sócrates” designa* um sujeito específico em uma frase como “Sócrates é mortal” se, e somente se, é esse sujeito específico que deve satisfazer o predicado “é mortal” para tornar a frase “Sócrates é mortal” verdadeira. Utilizando esse conceito de designação*, podemos definir os conceitos de expressão referencial modalmente rígida e modalmente flexível da seguinte maneira:

Uma expressão referencial “a” em uma frase como “Fa” é modalmente rígida se, e somente se, essa expressão “a” designa* o mesmo objeto x com respeito a qualquer mundo possível, i.e. é o mesmo objeto x que deve satisfazer o

predicado “F” para tornar a frase “Fa” verdadeira com respeito a qualquer mundo possível.

A ideia geral é que uma expressão referencial como “Sócrates” em uma frase como “Sócrates é mortal” é uma modalmente rígida se, e somente se, “Sócrates” designa* um mesmo sujeito com respeito a qualquer mundo possível, i.e. é o mesmo sujeito que deve satisfazer o predicado “é mortal” para tornar a frase “Sócrates é mortal” verdadeira com respeito a qualquer mundo possível. Para o caso das expressões referenciais modalmente flexíveis, a definição é a seguinte:

Uma expressão referencial “a” em uma frase como “Fa” é modalmente flexível se, e somente se, essa expressão “a” designa* diferentes objetos com respeito a diferentes mundos possíveis, i.e. são diferentes objetos que devem satisfazer o predicado “F” para tornar a frase “Fa” verdadeira com respeito a diferentes mundos possíveis.

A ideia geral é que uma expressão referencial como “o professor de Platão” em uma frase como “o professor de Platão é mortal” é uma expressão referencial modalmente flexível se, e somente se, “o professor de Platão” designa* diferentes sujeitos com respeito a diferentes mundos possíveis, i.e. são diferentes sujeitos que devem satisfazer o predicado “é mortal” para tornar a frase “o professor de Platão é mortal” verdadeira com respeito a diferentes mundos possíveis. Para o caso das frases, a definição de designação* é a seguinte:

Uma frase “s” designa* uma proposição p em uma frase como “Fs” se, e somente se, é a proposição p que deve satisfazer o predicado “F” para tornar a frase “Fs” verdadeira.

A ideia geral é que uma frase como “A neve é branca” designa* uma proposição específica em uma frase como “Que a neve é branca é conhecido” se, e somente se, é essa proposição específica que deve satisfazer o predicado “é conhecido” para tornar a frase “Que a neve é branca é conhecido” verdadeira. Utilizando esse conceito de designação*, podemos definir os conceitos de frase modalmente rígida e modalmente flexível da seguinte maneira:

Uma frase “s” em uma frase como “Fs” é modalmente rígida se, e somente se, essa frase “s” designa* a mesma proposição p com respeito a qualquer mundo possível, i.e. é a mesma proposição p que deve satisfazer o predicado “F” para tornar a frase “Fs” verdadeira com respeito a qualquer mundo possível.

A ideia geral é que uma frase como “A neve é branca” em uma frase como “Que a neve é branca é conhecido” é uma frase modalmente rígida se, e somente se, essa frase “A neve é branca” designa* uma mesma proposição com respeito a qualquer mundo possível, i.e. é a mesma proposição que deve satisfazer o predicado “é conhecido” para tornar a frase “Que a neve é branca é conhecido” verdadeira com respeito a qualquer mundo possível. Para o caso das frases modalmente flexíveis, a definição é a seguinte:

Uma frase “s” em uma frase como “Fs” é modalmente flexível se, e somente se, essa expressão “s” designa* diferentes proposições com respeito a diferentes mundos possíveis, i.e. são diferentes proposições que devem satisfazer o predicado “F” para tornar a frase “Fs” verdadeira com respeito a diferentes mundos possíveis.

A ideia geral é que uma frase como “A neve é branca” em uma frase como “Que a neve é branca é conhecido” é uma frase modalmente flexível se, e somente se, essa frase “A neve é branca” designa* diferentes proposições com respeito a diferentes mundos possíveis, i.e. são diferentes proposições que devem satisfazer o predicado “é conhecido” para tornar a frase “Que a neve é branca é conhecido” verdadeira com respeito a diferentes mundos possíveis.

Se essas definições estiverem em ordem, e penso que estejam, então faz sentido perguntar se a expressão “ $p_1 \wedge \neg Kp_1$ ” é uma expressão modalmente rígida ou uma expressão modalmente flexível.

2.3. Os aspectos centrais do argumento de Kvanvig

De acordo com Kvanvig, o mesmo tipo de invalidade que é produzido pelas expressões referenciais modalmente flexíveis também é produzida pelas frases modalmente flexíveis. Mais especificamente, assim como a substituição de expressões referenciais modalmente flexíveis no âmbito de operadores modais produzem alguns tipos de inferências inválidas, a substituição de frases modalmente flexíveis no âmbito de operadores modais também produz inferências inválidas.

Considere novamente nossa inferência problemática sobre o número de planetas. Como a expressão “o número de planetas” é uma expressão modalmente flexível, então, sob a leitura *de dicto*, essa inferência é inválida. Pois como a expressão “o número de planetas” é modalmente flexível, existe a possibilidade dessa expressão designar um número com respeito ao mundo atual e outro número com respeito a outros mundos possíveis.

Agora considere a seguinte inferência presente na demonstração do paradoxo da concecibilidade:

- b.i. Para qualquer proposição p , se a proposição p é verdadeira, então é possível que a proposição p seja conhecida;
- b.ii. Logo, se a proposição « p_1 é verdadeira e desconhecida» é verdadeira, então é possível que a proposição « p_1 é verdadeira e desconhecida» seja conhecida.

Se a expressão “ p_1 é verdadeira e desconhecida” for uma expressão modalmente flexível, então, sob a leitura *de dicto*, essa inferência será inválida. Pois se a expressão “ p_1 é verdadeira e desconhecida” for modalmente flexível, existe a possibilidade dessa expressão designar uma proposição com respeito ao mundo atual e outra proposição com respeito a outros mundos possíveis.

Essa é a ideia central da primeira premissa do argumento de Kvanvig, *viz.* que *todas as inferências que dependem da substituição de expressões modalmente flexíveis no âmbito de operadores modais são inferências inválidas.*

Agora que os pontos centrais da posição de Kvanvig estão devidamente esclarecidos, podemos passar para uma discussão mais crítica.

2.4. Uma objeção contra Kvanvig: a leitura *de re*

A discussão sobre as falácias modais na seção anterior nos fornece motivos para pensar que uma inferência presente no paradoxo da conhecibilidade é inválida sob a leitura *de dicto*. No entanto, Kvanvig não oferece motivo algum para pensar que a inferência presente no paradoxo é inválida sob a leitura *de re*. Ao contrário, nossa discussão sobre falácias modais nos fornece motivos para pensar que esse tipo de inferência é na verdade válida sob a leitura *de re*. Considere novamente nossa inferência sobre o número de planetas:

- (a.i) Para qualquer número n , se n é par, então é necessário que n seja par;
- (a.ii) Logo, se o número de planetas é par, então é necessário que o número de planetas seja par.

A premissa (a.i) atribui uma propriedade diretamente aos números, *viz.* a propriedade de necessariamente ser par, se atualmente for par. Sob a leitura *de re*, a conclusão (a.ii) também atribui essa propriedade diretamente ao número atualmente designado pela expressão “o número de planetas”, *viz.* a propriedade de necessariamente ser par, se atualmente for par. Nesse caso, como vimos, se a premissa for verdadeira, a conclusão também será verdadeira. Pois como essas propriedades estão sendo atribuídas diretamente aos números, não existe a possibilidade de estarmos falando de um número com respeito a um mundo possível, e de outro número com respeito a outro mundo possível. Por isso, nem todas as inferências que dependem da substituição de expressões modalmente flexíveis no âmbito de operadores modais são inválidas, *viz.* as leituras *de re* desse tipo de inferência são válidas.

Kvanvig poderia responder a essa objeção defendendo que apenas uma versão mais restrita dessa premissa é necessária para seu argumento. Ao invés de se comprometer com a afirmação mais geral de que qualquer expressão modalmente flexível sempre produzirá inferências inválidas, ele poderia se comprometer apenas com a afirmação mais restrita de que especificamente *frases* modalmente flexíveis sempre produzem inferências inválidas. Kvanvig poderia motivar essa ideia defendendo que, enquanto expressões referenciais modalmente flexíveis exibem uma ambiguidade entre uma leitura *de dicto* e uma leitura *de re*, frases modalmente flexíveis permitem apenas uma leitura *de dicto*. Assim, como a substituição de expressões modalmente flexíveis no âmbito de operadores modais são sempre inválidas sob a leitura *de dicto*, então todas as inferências que dependem da substituição de frases modalmente

flexíveis no âmbito de operadores modais seriam inválidas.

Mas isso também não funciona. Pois uma propriedade característica de expressões modalmente flexíveis, incluindo frases modalmente flexíveis, é que elas sempre produzem uma ambiguidade entre uma leitura *de dicto* e uma leitura *de re*. Suponha que a frase “ p_1 é verdadeira e desconhecida” seja uma frase modalmente flexível, e considere uma frase como “É possível que « p_1 é verdadeira e desconhecida» seja conhecida”. Por um lado, claramente podemos interpretar essa frase sob uma leitura *de dicto*, segundo a qual estamos atribuindo uma propriedade apenas ao *dicto* « p_1 é verdadeira e desconhecida», viz. a propriedade de possivelmente designar uma proposição conhecida, se atualmente designar uma proposição verdadeira. Mas, por outro lado, também claramente podemos interpretar essa frase sob uma leitura *de re*, segundo a qual estamos atribuindo uma propriedade diretamente à proposição atualmente designada pela frase « p_1 é verdadeira e desconhecida», viz. a propriedade de possivelmente ser conhecida, caso atualmente seja verdadeira.

Nesse caso, mesmo que a expressão “ p_1 é verdadeira e desconhecida” seja modalmente flexível, a demonstração de nosso paradoxo será válida se construída sob a leitura *de re*. Considere novamente a inferência em questão de nosso paradoxo:

- b.i. Para qualquer proposição p , se a proposição p é verdadeira, então é possível que a proposição p seja conhecida;
- b.ii. Logo, se a proposição p_1 é verdadeira e desconhecida é verdadeira, então é possível que a proposição p_1 é verdadeira e desconhecida seja conhecida.

Nesse caso, se a premissa for verdadeira, a conclusão também será verdadeira. Pois como as propriedades estão sendo atribuídas diretamente às proposições, então não existe a possibilidade de estarmos falando de uma proposição com respeito a um mundo possível, e de outra proposição com respeito a outro mundo possível.

Por isso, mesmo uma versão mais restrita dessa premissa continua sendo falsa. Nem todas as inferências que dependem da substituição de frases modalmente flexíveis, ou mais especificamente da expressão “ p_1 é verdadeira e desconhecida”, no âmbito de operadores modais são inválidas, viz. as leituras *de re* dessas inferências são válidas.

2.5. Outra objeção contra Kvanvig: a necessidade fixa

Até agora argumentei que existe uma leitura *de re* na qual a demonstração envolvida em nosso paradoxo é válida. Minha segunda objeção é que, mesmo que existisse apenas a leitura *de dicto*, ainda seria possível reconstruir essa demonstração utilizando do um conceito diferente de modalidade.

A ideia de definir diferentes conceitos de modalidade é uma ideia mais ou menos familiar. Como mencionado, nossa noção usual de modalidade pode ser representada como uma quantificação sobre mundos possíveis. Nossa noção usual de necessidade é representada como uma quantificação universal sobre mundos possíveis: algo é necessário se, e somente se, é o caso com respeito a todos os mundos possíveis. Similarmente, nossa noção usual de possibilidade é representada como uma quantificação existencial sobre mundos possíveis: algo é possível se, e somente se, é o caso com respeito a algum mundo possível.

Além de nossa noção usual de modalidade, podemos definir uma noção de modalidade física. Essa noção de modalidade física também pode ser representada como uma quantificação sobre mundos possíveis, com a diferença que restringimos essa quantificação para incluir apenas os mundos possíveis que possuam as mesmas leis da física que o mundo atual. Nesse caso, algo é fisicamente necessário se, e somente se, é o caso com respeito a todos os mundos possíveis que possuam as mesmas leis da física que o mundo atual. Similarmente, algo é fisicamente possível se, e somente se, é o caso com respeito a algum mundo possível que possua as mesmas leis da física que o mundo atual.

Em adição à noção de modalidade física, podemos também definir uma noção de *modalidade fixa*. Nesse caso restringimos a quantificação para incluir apenas os mundos possíveis que possuam o mesmo domínio de quantificação que o mundo atual. Assim, algo é fixamente necessário se, e somente se, é o caso com respeito a todos os mundos possíveis que possuam o mesmo domínio de quantificação que o mundo atual. Similarmente, algo é fixamente possível se, e somente se, é o caso com respeito a algum mundo possível que possua o mesmo domínio de quantificação que o mundo atual.

A consequência dessa noção de necessidade fixa é que, mesmo aceitando que a expressão “ p_1 é verdadeira e desconhecida” é modalmente flexível na noção usual de modalidade, ela não será modalmente flexível nessa nova noção de modalidade. Essa definição bloqueia a solução de Kvanvig porque ele defende que a expressão “ p_1 é verdadeira e desconhecida” é modalmente flexível em virtude dessa expressão ser a abreviação da expressão quantificada “ p_1 é verdadeira e não existe um sujeito que a conheça”, e em virtude do domínio de quantificação

ser um elemento constituinte da proposição expressa por expressões quantificadas. E uma consequência dessa posição é que a expressão “ p_1 é verdadeira e desconhecida” é modalmente flexível apenas se diferentes mundos possíveis possuírem diferentes domínios de quantificação.

Assim, se utilizarmos essa noção de necessidade fixa, então a expressão “ p_1 é verdadeira e desconhecida” será modalmente rígida, e poderemos utilizar essa noção para reconstruir a inferência envolvida no paradoxo da conhecibilidade:

- b.i. Para qualquer proposição p , se a proposição p é verdadeira, então é fixamente possível que a proposição p seja conhecida;
- b.ii. Logo, se a proposição p_1 é verdadeira e desconhecida é verdadeira, então é fixamente possível que a proposição p_1 é verdadeira e desconhecida seja conhecida.

Como nossa noção de necessidade fixa exclui a variação no domínio de quantificação de mundo possível para mundo possível, então a expressão “ p_1 é verdadeira e desconhecida” será modalmente rígida. Isso quer dizer que nossa demonstração será garantidamente válida, independentemente de ser construída sob a leitura *de dicto* ou sob a leitura *de re*.

2.6. Um pouco sobre a segunda premissa do argumento de Kvanvig

Até agora discutimos mais detalhadamente a premissa A.1 do argumento de Kvanvig, i.e. a premissa segundo a qual *todas as inferências que dependem da substituição de expressões modalmente flexíveis no âmbito de operadores modais são inferências inválidas*. Argumentei que essa premissa é falsa: A leitura *de re* dessas inferências são válidas e é possível definir conceitos de necessidade e possibilidade segundo os quais esse tipo de inferência é válida nas duas leituras. Portanto, penso que Kvanvig falha em defender que a demonstração do paradoxo da conhecibilidade é inválida.

Em todo caso, é interessante avaliar também a premissa A.2 do argumento de Kvanvig, i.e. a premissa segundo a qual *uma das inferências da demonstração do paradoxo da conhecibilidade depende da substituição de uma expressão modalmente flexível no âmbito de um operador modal*. A ideia central dessa premissa é que a expressão “ $p_1 \wedge \neg Kp_1$ ” é modalmente

flexível. Em linhas gerais, Kvanvig defende essa ideia com o seguinte argumento:

- B.1. A teoria semântica correta das frases quantificadas é a teoria neo-russelliana da quantificação;
- B.2. Se a teoria neo-russelliana da quantificação é correta, então expressões quantificadas são modalmente indexicais;
- B.3. Expressões modalmente indexicais são expressões modalmente flexíveis;
- B.4. Logo, como a expressão " $p_1 \wedge \neg Kp_1$ " é a abreviação da expressão quantificada " $p_1 \wedge \neg \exists s \exists t Kstp_1$ ", então essa expressão é modalmente flexível.

Uma discussão mais detalhada desses pontos foge ao escopo deste artigo, mas o leitor interessado pode consultar Francisco (2013), onde discuto esses pontos em mais detalhes.

3. Conclusão

Como o leitor provavelmente percebeu, o paradoxo da conhecibilidade é um problema relativamente complexo, que se relaciona com problemas de diferentes áreas da filosofia.

Discussões na área da filosofia da linguagem são relevantes para sabermos se alguma inferência na demonstração do paradoxo depende de alguma substituição de expressões modalmente flexíveis no âmbito de operadores modais, e até que ponto isso é problemático. Discussões na área da filosofia da lógica são relevantes para sabermos se, dentre outros, o conceito de conhecimento de diferentes ordens devem ser distinguidos, ou se contradições podem ser verdadeiras, ou se a lógica intuicionista é preferível à lógica clássica. Finalmente, discussões na área da teoria do conhecimento são relevantes para entendermos melhor a natureza do conceito de conhecibilidade, e considerarmos se a tese da conhecibilidade deve ser formulada em termos de conhecimento possível, ou capacidades para conhecer, ou conhecimento potencial.

Assim como outras áreas da filosofia são relevantes para o paradoxo da conhecibilidade, os problemas levantados por esse paradoxo também prometem ser relevantes para outras áreas da filosofia. O exemplo mais direto aqui, sem dúvida, são as teorias anti-realistas afetadas

diretamente pelo paradoxo. Independentemente de ele ser uma refutação dessas teorias ou de ser uma motivação para encontrarmos formulações mais sofisticadas delas, o paradoxo da conhecibilidade sem dúvida força os anti-realistas a confrontarem novos problemas.

Portanto, uma compreensão aprofundada do paradoxo da conhecibilidade promete uma compreensão correspondentemente aprofundada sobre diversas questões cruciais sobre os limites de nosso conhecimento e sobre diversos outros problemas em diferentes áreas da filosofia.

Referências

Beall, J. C. "Knowability and Possible Epistemic Oddities", in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009 , 106-125

Beall, J. C. "Fitch's proof, verificationism, and the knower paradox". *Australasian Journal of Philosophy*, 2000 , 78 , 241-247

Bell, D. & Hart, W. D. "The Epistemology of Abstract Objects". *Proceedings of the Aristotelian Society, Supplementary Volumes*, 1979 , 53 , 1-32

Bermudez, J. L. "Truth, Indefinite Extensibility, and Fitch's Paradox" in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009

Brogaard, B. & Salerno, J. "Clues to the paradoxes of knowability: reply to Dummett and Tennant". *Analysis*, Oxford University Press, 2002 , 62 , 143-150

Brogaard, B. & Salerno, J. "Knowability and a Modal Closure Principle", *American Philosophical Quarterly*, University of Illinois Press, 2006 , 43 , 261-270

Brogaard, B. & Salerno, J. "Knowability, Possibility and Paradox", *New Waves in Epistemology*, editado por V. Hendricks, Palgrave Macmillan, 2008

Carrara, M. & Fassio, D. "Why Knowledge Should Not Be Typed: An Argument against the Type Solution to the Knowability Paradox". *Theoria*, Philosophy Department, Stockholm University, 2011 , 77 , 180-193

Church, A. "Referee Reports on Fitch's 'A Definition of Value' ", in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009 , 13-20

DeVidi, D. & Solomon, G. "Knowability and Intuitionistic Logic", *Philosophia*, Springer

Netherlands, 2001 , 28 , 319-334

DeVidi, D. & Kenyon, T. "Analogues of Knowability". Australasian Journal of Philosophy, Taylor & Francis, 2003 , 81 , 481-495

Dummett, M. "Fitch's Paradox of Knowability" in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009 , 51-52

Edgington, D. "The Paradox of Knowability". *Mind*, 1985 , 94 , 557-568

Edgington, D. "Possible knowledge of unknown truth". *Synthese*, 2010 , 173 , 41-52

Fara, M. "Knowability and the capacity to know". *Synthese*, 2010 , 173 , 53-73

Fitch, F. B. "A Logical Analysis of Some Value Concepts". *The Journal of Symbolic Logic*, 1963 , 28 , 135-142

Francisco, I. B. *Será o paradoxo da conhecibilidade uma falácia modal? Dissertação (Mestrado em Filosofia) – Programa de Pós Graduação em Lógica e Metafísica, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2013.*

Fuhrmann, A. "Knowability as Potential Knowledge". *Synthese*, 2014 , 191 , 1627-1648

Halbach, V. "On a side effect of solving Fitch's paradox by typing knowledge". *Analysis*, Oxford University Press, 2008 , 68 , 114-120

Hart, W. D. & McGinn, C. "Knowledge and necessity". *Journal of Philosophical Logic*, Springer, 1976 , 5 , 205-208

Hand, M. & Kvanvig, J. L. "Tennant on Knowability". *Australasian Journal of Philosophy*, 1999 , 77 , 422-428

Hand, M. "Knowability and Epistemic Truth". *Australasian Journal of Philosophy*, 2003 , 81 , 216-228

Kvanvig, J. L. "The Knowability Paradox and the Prospects for Anti-Realism". *Noûs*, 1996 , 29 , 481-500

Kvanvig, J. L. *The Knowability Paradox*. Oxford University Press, 2006

Lindström, S. "Situations, Truth and Knowability: A Situation-Theoretic Analysis of a Paradox by Fitch" in *Logic, Action and Cognition: Essays in Philosophical Logic*, editado por E. Ejerhed & S. Lindström, Kluwer Academic Publishers, 1997 , 183-210

Linsky, B. "Logical Types in Some Arguments about Knowability and Belief", in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009, 163-181

- Mackie, J. L. "Truth and knowability". *Analysis*, 1980 , 40 , 90-92
- Paseau, A. "Fitch's argument and typing knowledge". *Notre Dame Journal of Formal Logic*, University of Notre Dame, 2008 , 49 , 153-176
- Maffezioli, P.; Naibo, A. & Negri, S. "The Church-Fitch knowability paradox in the light of structural proof theory". *Synthese*, 2012
- Paseau, A. "How to type: reply to Halbach". *Analysis*, Oxford University Press, 2009 , 69 , 280-286
- Percival, P. "Knowability, actuality, and the metaphysics of context-dependence", *Australasian Journal of Philosophy*, 1991 , 69 , 37-41
- Percival, P. "Fitch and Intuitionistic Knowability", *Analysis*, Oxford University Press, 1990 , 50 , 182-187
- Priest, G. "Beyond the Limits of Knowledge" in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009 , 93-104
- Rabinowicz, W. & Segerberg, K. "Actual truth, possible knowledge". *Topoi*, 1994 , 13 , 101-115
- Rasmussen, S. A. "The Paradox of Knowability and the Mapping Objection" in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009 , 53-75
- Routley, R. "Necessary Limits to Knowledge: Unknowable Truths" in *Essays in Scientific Philosophy*, Dedicated to Paul Weingartner / Philosophie als Wissenschaft, Paul Weingartner gewidmet, editado por M. Edgar N. Otto & Z. Gerhard, Bad Reichenhall: Comes Verlag, 1981 , 93-115. (Reimpresso em *Synthese*, 2010 , 173 , 107-122)
- Rosenkranz, S. "Fitch back in action again?". *Analysis*, Blackwell Synergy, 2004 , 64 , 67-71
- Schlesinger, G. N. *The Range of Epistemic Logic*. Aberdeen University Press, 1985
- Tennant, N. "Is every truth knowable? Reply to Hand and Kvanvig". *Australasian journal of philosophy*, 2001 , 79 , 107-113
- Tennant, N. "Is every truth knowable? Reply to Williamson". *Ratio*, 2001 , 14 , 263-280
- Tennant, N. "Victor vanquished" *Analysis*, Oxford University Press, 2002 , 62 , 135-142
- Tennant, N. "Revamping the Restriction Strategy" in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009 , 224-238
- Tennant, N. "Williamson's Woes". *Synthese*, 2010 , 173 , 9-23

- Williamson, T. "Intuitionism disproved?". *Analysis*, 1982 , 42 , 203-207
- Williamson, T. "On Knowledge of the Unknowable". *Analysis*, 1987 , 47 , 154-158
- Williamson, T. "Knowability and constructivism" *The Philosophical Quarterly*, 1988 , 38 , 422-432
- Williamson, T. "On intuitionistic modal epistemic logic", *Journal of Philosophical Logic*, 1992 , 21 , 63-89
- Williamson, T. "Structural Unknowability", in *Knowledge and its Limits*, Capítulo 12, Oxford University Press, 2000 , 270-301
- Williamson, T. "Tennant on knowable truth". *Ratio*, 2000 , 13 , 99-114
- Williamson, T. "Tennant's Troubles" in *New Essays on the Knowability Paradox*, editado por Joe Salerno, Oxford University Press, 2009 , 183-204
- Williamson, T. "Edgington on Possible Knowledge of Unknown Truth" in *Conditionals, Probability & Paradox: Themes from the Philosophy of Dorothy Edgington*, Oxford University Press, no prelo.

Algumas considerações sobre o Paradoxo de Curry

Some remarks on Curry's Paradox

André Nascimento Pontes
Universidade Federal do Amazonas

Resumo

O presente artigo tem por objetivo realizar uma breve discussão sobre o paradoxo de Curry e sua propriedade de ser livre da negação (*negation-free*). Além disso, discorro também sobre a correlação do paradoxo em questão com importantes tópicos da filosofia da lógica, tais como os conceitos de consistência e verdade.

Palavras-chave

Curry. Paradoxo. Negação.

Abstract

The purpose of this paper is to hold a short discussion about Curry's paradox and its property of being negation-free. In addition, I also go into the correlation of the paradox at hand with important topics on the philosophy of logic, such as the concepts of consistency and truth.

Key-words

Curry. Paradox. Negation.

Introdução

De um ponto de vista estritamente formal, um paradoxo é caracterizado como um argumento válido contendo premissas amplamente aceitas como verdadeiras, mas dotado de uma conclusão falsa. Os paradoxos produzem grande desconforto na medida em que, sendo argumentos válidos, eles deveriam ser capazes de preservar verdade. Afinal, nosso entendimento padrão de validade nos diz que um argumento é válido quando a pura forma lógica do argumento garante que, caso as premissas sejam verdadeiras, a conclusão também será. No entanto, de maneira desconcertante, partindo de sentenças aceitas como verdadeiras, as conclusões obtidas em argumentos paradoxais são falsas. Essas conclusões falsas podem figurar de três modos básicos. Elas podem ser:

- (i) Uma contradição explícita, ou seja, uma sentença do tipo $p \wedge \neg p$;
- (ii) Uma contradição implícita, ou seja, uma sentença que não é *per se* uma contradição, mas que, no contexto da teoria em que ela está inserida, possui uma carga inferencial contraditória;

ou ainda;

- (iii) Uma sentença cuja suposta verdade seria fortemente contraintuitiva.

Exemplos de paradoxos que ilustram as conclusões dos tipos (i), (ii) e (iii) são, respectivamente, os paradoxos de Russell, de Curry e a classe de paradoxos conhecida como *Sorites*.

Costumeiramente, os chamados paradoxos lógicos são dotados de conclusões do tipo (i) e, portanto, envolvem a noção clássica de negação. Mas isso nem sempre é o caso. Publicado em 1942 pelo lógico e matemático estadunidense Haskell B. Curry, o chamado *paradoxo de Curry* é mais um dos resultados que compõem a classe dos chamados paradoxos da auto-referência, ou paradoxos da circularidade – tais como o paradoxo de Russell, o paradoxo do mentiroso, o paradoxo de Grelling-Nelson, dentre tantos outros –, mas que, ao contrário dos paradoxos mencionados, dispensa o uso de negação (*negation-free*).

Alguns autores, tais com Haack (2002: p.190), sustentam uma distinção entre uma *solução formal* e uma *solução filosófica* aos argumentos paradoxais. A solução formal consiste em indicar quais premissas ou princípios aparentemente verdadeiros e utilizados no argumento são, na verdade, problemáticos e que, por isso, devem ser eliminados. Já uma solução filosófica, em princípio, deveria oferecer uma razão extra lógica que fundamente a rejeição de alguma premissa ou princípio do argumento em questão. Essa razão é dita *extra lógica*, pois ela deve ter um caráter independente do simples fato de que a premissa ou princípio esteja envolvido em um argumento paradoxal.

Se realmente levarmos em consideração a possibilidade de que haja uma solução única que evite o surgimento de paradoxos, seja de que tipo eles forem, devemos também assumir a necessidade de que tal solução satisfaça a dois critérios básicos: (i) ela não pode ser demasiadamente forte para desabilitar não só os argumentos paradoxais, mas também argumentos não paradoxais; (ii) ela não pode ser demasiadamente fraca ao ponto de que haja algum paradoxo que escape incólume às restrições impostas pela solução. Em resumo, uma solução única pressuposta por uma teoria geral de paradoxos deve satisfazer ao lema *in medio stat virtus*; ela deve ser um remédio administrado na justa medida.

No presente artigo não pretendo dar conta de uma teoria geral de paradoxos. Meu ponto é mais modesto. No que segue, meu objetivo será o de realizar uma apresentação do paradoxo de Curry, bem como de alguns desdobramentos lógicos e filosóficos que podem ser levantados a partir dele. O resultado obtido por Curry pode ser, *mutatis mutandis*, apresentado em duas versões. Uma versão é demonstrada com base em operações da teoria dos conjuntos (*set-theoretic*) e outra estabelecida dentro da teoria da verdade (*truth-theoretic*). Com isso, na primeira seção, tento reconstruir a demonstração do resultado obtido

por Curry em contraste com o famigerado paradoxo de Russell. Nesta mesma seção, mostro como o paradoxo de Curry dispensa o apelo ao operador de negação. Essa apresentação do paradoxo de Curry é realizada no seio da teoria ingênua dos conjuntos. Na segunda seção, apresento a relação do resultado de Curry com a concepção ingênua de verdade. Neste contexto, há uma estreita relação entre o resultado de Curry e o tão celebrado paradoxo do mentiroso na medida em que ambos revelam que uma linguagem que contenha seu próprio predicado de verdade – o que Tarski chamou de *linguagem semanticamente fechada* – é uma linguagem inconsistente.

1. Russell e Curry: dois resultados em torno do princípio ingênuo da compreensão

Juntamente com o paradoxo de Russell, o resultado obtido por Curry é recorrentemente compreendido como prova da inconsistência da teoria ingênua dos conjuntos. No entanto, o paradoxo de Russell é certamente mais famoso na medida em que ele foi obtido num contexto mais amplo de prova da inconsistência do programa logicista fregeano. Nas duas próximas seções apresento os traços gerais dos dois resultados e tento destacar semelhanças e diferença entre eles.

1.1. O paradoxo de Russell

Há basicamente duas formas de apresentar um conjunto, a saber, uma *extensional* e outra *intensional*. Quando apresentamos um conjunto através de uma lista exaustiva de todos os elementos que pertencem a ele, apresentamos uma definição extensional do conjunto em questão. Por outro lado, se expressamos o conjunto através de uma propriedade ou condição comum compartilhada por todos os elementos do conjunto em questão, estamos oferecendo uma definição intensional do conjunto. Por exemplo, o conjunto C de seleções campeãs mundiais de futebol pode ser expresso extensionalmente por:

$$B = \{\text{Brasil, Uruguai, Argentina, Alemanha, Itália, França, Espanha, Inglaterra}\}$$

ou intensionalmente por

$$B = \{x \mid x \text{ é uma seleção campeã mundial de futebol}\}$$

(Lê-se: “B é o conjunto de elementos x , tal que x satisfaz a condição *ser seleção campeã mundial de futebol*” ou simplesmente, “B é o conjunto de elementos x , tal que x é uma seleção campeã mundial de futebol”)

No caso de conjuntos apresentados extensionalmente, a ordem em que os elementos ocorrem na lista não importa e repetições são permitidas. Sendo assim, os três conjuntos abaixo são exatamente o mesmo conjunto.

$$A = \{1, 2, 3, 4\}$$

$$A' = \{2, 4, 3, 1\}$$

$$A'' = \{1, 2, 3, 4, 2\}$$

Quanto ao modo intensional, ele é costumeiramente expresso em livros texto de lógica e matemática básica quando conjuntos são introduzidos como uma coleção ou agregado de objetos distintos selecionados por uma lei.

Na origem histórica da teoria dos conjuntos, um importante princípio foi proposto de modo a correlacionar intensões e extensões. Amplamente conhecido como *princípio ingênuo da compreensão*, ele foi formulado em uma versão ainda não axiomatizada da teoria e, posteriormente, se mostrou inconsistente, pois por intermédio dele é possível derivar uma contradição (veja abaixo a demonstração do *paradoxo de Russell*). A ideia básica expressa no princípio em questão é a de que para toda propriedade ou condição Φ , há um conjunto β que possui como elementos apenas os objetos que satisfazem Φ . Em termos formais temos:

Princípio Ingênuo da Compreensão

$$\forall x \forall \Phi \exists \beta (x \in \beta \leftrightarrow \Phi(x))$$

Em outras palavras, toda propriedade ou condição Φ – sendo aqui Φ uma intensão – determina um conjunto β formado pelos objetos que exemplificam Φ . É precisamente a esse conjunto β que chamamos de extensão de Φ .

A relação entre intensão e extensão pode ser expressa no seguinte esquema:

$$\varphi(x) \mapsto \{x \mid \varphi(x)\}$$

onde do lado esquerdo temos uma dada propriedade Φ e do lado direito o conjunto de objetos de que satisfaz a propriedade em questão. Por exemplo, parece facilmente aceitável que se escolhermos arbitrariamente uma propriedade, por exemplo, *ser azul*, associado a ela temos um conjunto de todos os objetos azuis. Isso valeria também para propriedades que não são satisfeitas por nada. Por exemplo, a propriedade de *ser o maior número primo* possui uma extensão nula e, portanto, a ela está associado o conjunto vazio.

Embora bastante intuitivo, o princípio ingênuo da compreensão foi irremediavelmente abalado por um resultado obtido por Bertrand Russell. Na prática, o que Russell mostrou foi que o princípio em questão não é universalmente aplicável. Nem toda intensão corresponde a uma extensão. Para isso, Russell tomou uma propriedade ou condição definida dentro da própria teoria dos conjuntos, a saber, *conjunto que não pertence a si mesmo* e mostrou, por redução ao absurdo, que não existe uma extensão ou conjunto associado a ela. Desse modo, Russell mostrou que o princípio ingênuo da compreensão leva à inconsistência a teoria dos conjuntos que o contém, pois podemos obter uma contradição a partir de tal princípio. Uma pequena reconstrução da demonstração pode ser dada como segue.

Definição: $R = \{x \mid x \notin x\}$

- (1) $\forall x \forall \Phi \exists \beta (x \in \beta \leftrightarrow \Phi(x))$ [princípio ingênuo da compreensão]
- (2) $\forall x (x \in R \leftrightarrow x \notin x)$ [instanciação universal e existencial de (1) para o caso R]
- (3) $R \in R \leftrightarrow R \notin R$ ■ [instanciação universal de (2) para o caso R]

A definição introduz um conjunto R [de Russell] que contém todos os conjuntos que não são membros de si mesmo ($x \notin x$). Esse parece ser um conjunto perfeitamente legítimo, uma vez que todo conjunto ou é ou não é membro de si mesmo. Por exemplo, o conjunto B de *seleções campeãs mundiais de futebol*, não é membro de si mesmo, pois ele não satisfaz a condição que o determina; afinal, B é um conjunto e não uma seleção de futebol. Já o conjunto de todos os conjuntos com mais de três elementos é membro de si mesmo, uma vez que ele satisfaz a condição que o determina; ele possui certamente mais de três elementos.

Com base no princípio ingênuo da compreensão, a intensão ou condição *não ser membro de si mesmo* ($x \notin x$) deveria determinar uma extensão definida aqui como o conjunto R . No entanto, dado que R , ele próprio, é um conjunto, podemos perguntar: R é ou não um membro de si mesmo? A passagem de (2) a (3) revela que essa questão leva a uma contradição e que, portanto, não há nenhum conjunto R associado à condição *não ser membro de si mesmo*. Em (3) temos claramente uma contradição! (3) é uma contradição explícita e isso fica evidente uma vez que ela é equivalente a

$$(3^*) R \in R \wedge R \notin R$$

Logo, o resultado de Russell revela que o princípio ingênuo da compreensão não pode ser consistentemente endossado. Caso contrário, ele traria para dentro da teoria dos conjuntos que o admitisse uma contradição e, por conseguinte, a inconsistência e o trivialismo dedutivo da teoria em questão. Mais adiante irei retomar esse ponto do trivialismo dedutivo de teorias inconsistentes.

Vale destacar ainda que o resultado obtido por Russell depende fundamentalmente do operador lógico da negação em pelo menos dois aspectos: (i) a condição de partida do paradoxo é a propriedade ($x \notin x$) que nada mais é que a negação da relação de pertença, ou seja, $\neg(x \in x)$; (ii) a conclusão do argumento, na medida em que ela é uma contradição explícita, também envolve uma negação.

$$(3^{**}) R \in R \wedge \neg(R \in R)$$

E isso é o básico que precisamos saber sobre o paradoxo de Russell para os objetivos do presente artigo.

1.2. O Paradoxo de Curry

Antes de apresentar o paradoxo de Curry, vale introduzir o leitor iniciante em uma bastante útil ferramenta da lógica clássica conhecida como *Lei de Contração*:¹

Lei da Contração

$$(\alpha \rightarrow (\alpha \rightarrow \beta)) \rightarrow (\alpha \rightarrow \beta)$$

Ela afirma basicamente que se uma sentença α implica uma sentença do tipo $\alpha \rightarrow \beta$ – onde α é, ela mesma, o antecedente –, então α implica o conseqüente dessa mesma implicação, ou seja, $\alpha \rightarrow \beta$. Esse é um dos mecanismos disponíveis em lógica clássica que simplificam o raciocínio apresentando uma sentença mais elementar como equivalente a uma sentença mais complexa. Voltemos agora nossa atenção para a interessante descoberta de Curry.

O resultado apresentado por Curry tem a interessante característica de ser obtido por intermédio da substituição da condição que Russell propôs ($x \notin x$), pela condição $x \in x \rightarrow B$, onde não há ocorrência do operador de negação e B figura como uma sentença arbitrária. Desse modo, dado o esquema geral:²

$$(4) \exists y \forall x (x \in y \leftrightarrow Fx)$$

Vale aqui destacar que (4) expressa uma forma concisa do princípio da compreensão apresentado anteriormente. Ao substituir em (4) a variável predicativa Fx pela condição $x \in x \rightarrow B$ obtemos a seguinte sentença:

1 - Uma demonstração de que a Lei de Contração é de fato uma verdade lógica – ou seja, uma tautologia – pode ser obtida ao se construir a tabela de verdade para a sentença que a expressa, ou ainda, pela construção de uma árvore semântica desta mesma sentença.

2 - A presente demonstração do paradoxo de Curry reproduz, em linhas gerais, a demonstração contida em Da Costa *et. al.* (1998). Uma versão bastante didática do paradoxo de Curry foi apresentada por Prior (1955).

$$(5) \exists y \forall x (x \in y \leftrightarrow x \in x \rightarrow B)$$

Assumindo c como o conjunto postulado pela variável y em (4), segue-se:

$$(6) \forall x (x \in c \leftrightarrow x \in x \rightarrow B)$$

Em seguida, ao instanciarmos a quantificação universal em (6) para o caso c , obtemos:

$$(7) c \in c \leftrightarrow (c \in c \rightarrow B)$$

Por eliminação do bicondicional em (7):

$$(8) c \in c \rightarrow (c \in c \rightarrow B)$$

e

$$(9) (c \in c \rightarrow B) \rightarrow c \in c$$

Por intermédio da Lei de Contração aplicada a (8) obtemos:

$$(10) c \in c \rightarrow B$$

e por *modus ponens* de (9) e (10):

$$(11) c \in c$$

e, por *modus ponens* de (10) e (11):

(12) B ■

Sendo B uma sentença arbitrária, como foi assumido acima, o paradoxo de Curry revela que toda teoria dos conjuntos que inclua o princípio da compreensão é trivialmente dedutiva, uma vez que ela pode obter uma derivação de toda sentença do sistema. Em outras palavras, toda sentença do sistema é também um teorema do sistema! Em termos clássicos, esse trivialismo dedutivo implica em inconsistência formal, uma vez que dada uma sentença p de um sistema formal S trivialmente dedutivo, ambas p e $\neg p$ seriam derivadas em S. Desse modo, o paradoxo de Curry funciona como uma prova da inconsistência do sistema onde ele foi obtido, pois mostra que podemos provar sentenças contraditórias no sistema em questão.

Um importante princípio clássico revela também um caminho inverso ao apresentado acima. O raciocínio descrito anteriormente mostra que, do ponto de vista clássico no qual o resultado de Curry foi obtido, se eu posso derivar toda sentença do sistema, então eu posso derivar sentenças contraditórias. O chamado princípio da explosão afirma que, caso haja uma sentença contraditória como um axioma ou teorema do meu sistema, então eu posso derivar toda sentença do sistema em questão; o que equivale ao trivialismo dedutivo do sistema.

Princípio da Explosão	
<p>Sendo $\alpha \wedge \neg \alpha$ um axioma de um dado sistema S, segue-se que:</p> $\alpha \wedge \neg \alpha \vdash_S \beta$	<p>Sendo α e $\neg \alpha$ sentenças de um dado sistema formal S, Γ um conjunto de sentenças e β uma sentença arbitrária do mesmo sistema S, segue-se que:</p> $\frac{\Gamma \vdash_S \alpha \quad \Gamma \vdash_S \neg \alpha}{\Gamma \vdash_S \beta}$

2. O paradoxo de Curry e as linguagens semanticamente fechadas

O paradoxo de Curry é também responsável por lançar um importante desafio à compreensão ingênua do conceito de verdade. Nessa abordagem do resultado de Curry, é possível perceber que uma concepção teórica que não imponha restrições ao uso do predicado de verdade – especialmente, àqueles usos que associam tal predicado a procedimentos envolvendo circularidade ou auto-referência – torna inconsistente a linguagem que comporta tal predicado. Vejamos de maneira bastante direta como isso ocorre.³

Se uma determinada abordagem da noção de verdade pode ser capturada pelo esquema-T do seguinte tipo

$$\text{(esquema-T)} \quad T[p] \leftrightarrow p$$

onde o operador $T[\dots]$ expressa uma atribuição de verdade a uma dada sentença e p é uma sentença arbitrária da linguagem para qual estamos definindo o conceito de verdade, segue-se que, caso não estabeleçamos nenhum tipo de restrição a aplicações circulares ou auto-referenciais do esquema-T, a concepção de verdade para a linguagem em questão será inconsistente. Essa conclusão pode ser demonstrada em um número relativamente pequeno de passos da seguinte maneira. Assuma a sentença auto-referencial L de uma dada linguagem \mathcal{L} semanticamente fechada que afirma de si mesma que se ela é verdadeira, então é o caso que toda sentença S de \mathcal{L} é verdadeira. Ou seja,

$$(L) \quad T[L] \rightarrow S$$

Curiosamente, sendo $(T[L] \rightarrow S)$ uma sentença de \mathcal{L} , podemos tomá-la como uma instância substitutiva do esquema-T acima apresentado e obter:

$$(13) \quad T[L] \leftrightarrow (T[L] \rightarrow S)$$

Por eliminação do bicondicional de (13) temos:

3 - A presente versão do paradoxo de Curry aplicado ao conceito de verdade é uma reconstrução da versão apresentada por Beall (2008).

$$(14) T[L] \rightarrow (T[L] \rightarrow S)$$

e

$$(15) (T[L] \rightarrow S) \rightarrow T[L]$$

Aplicando a *Lei de Contração* – cf. subseção 1.2. – a (14), segue-se que:

$$(16) T[L] \rightarrow S$$

Por *modus ponens* de (15) e (16):

$$(17) T[L]$$

Finalmente, por *modus ponens* de (L) e (17) segue-se que:

$$(18) S \quad \blacksquare$$

E novamente o paradoxo de Curry insiste em conduzir nossas teorias ao trivialismo dedutivo!

Nesse contexto, a principal estratégia de enfrentamento ao problema encontra-se na obra de Tarski.⁴ A célebre distinção de níveis de linguagem estabelecida por Tarski (1933; 1944) tem como uma de suas principais finalidades evitar a derivação de paradoxos obtidos quando usamos o esquema-T de forma irrestrita, a exemplo do que ocorre no caso dos paradoxos do mentiroso e de Curry em sua versão *truth-theoretic*. Tarski chamava de semanticamente fechadas as linguagens que permitem o uso irrestrito do esquema-T de modo que elas contêm o predicado de verdade que elas aplicam às suas próprias sentenças.

4 - Outras abordagens podem ser encontradas em Parsons (1974), Kripke (1975), Gupta-Belnap (1993) e Glanzberg (2001).

De modo contrário, as linguagens que restringem o uso do esquema-T baseado em uma hierarquia de sistemas linguísticos impondo a aplicação do predicado de verdade apenas às sentenças de uma linguagem hierarquicamente inferior eram chamadas por Tarski de *semanticamente abertas*. O que está por trás da hierarquia tarskiana de linguagens é, dentre outras coisas, bloquear o uso de circularidades como a presente na sentença (L) acima.

3. Algumas considerações finais

Dado o exposto acima, fica patente a estreita relação entre o paradoxo de Curry e uma vasta gama de tópicos de profundo interesse em lógica e filosofia, tais como a consistência de sistemas formais, a natureza da negação, a semântica do predicado de verdade, etc. Uma avaliação do resultado obtido por Curry tem certamente muito a revelar. Um exemplo paradigmático é dado pela consequência derivada do resultado de Curry de que não só a ocorrência de contradições explícitas não é necessária para o surgimento de paradoxos, como também não é necessário o operador clássico de negação. Nesse aspecto, é destacável o interesse de lógicos não clássicos com relação ao paradoxo de Curry. Tanto a lógica paraconsistente quanto a lógica paracompleta possuem abordagens do resultado de Curry que buscam evitar suas consequências mais radicais. Para compreensão do iniciante, a lógica paraconsistente busca, ao rejeitar algumas regras clássicas de inferência, barrar o princípio de explosão. Com isso, lógicas paraconsistentes se caracterizam por ser inconsistente, porém não dedutivamente triviais; algo que, do ponto de vista clássico, é inconcebível. Já a lógica paracompleta rejeita o princípio clássico do terceiro excluído, ou seja, a verdade da disjunção entre uma sentença P qualquer e sua negação. Ela sustenta que há sentenças P^* , tal que ela e sua negação sejam ambas falsas; o que tornaria falso $P^* \vee \neg P^*$.

Outra característica que durante muitos anos foi considerada por filósofos, lógicos e matemáticos como um constituinte fundamental dos paradoxos foi a circularidade/impredicatividade. Russell e Poincaré fizeram defesas bastante enfáticas de que nossas teorias deveriam evitar procedimentos que envolvessem circularidade e definições impredicativas se quisessem evitar o surgimento de paradoxos. No entanto, mesmo essa proposta mostrou-se posteriormente insuficiente. O chamado paradoxo linear de Yablo (1993) revelou que circularidade e impredicatividade não são propriedades necessárias de paradoxos.

Desse modo, penso que, juntamente com o paradoxo linear de Yablo, o resultado obtido por Curry representou o calcanhar de Aquiles das principais propostas de teoria geral de paradoxos que pretendiam eliminar argumentos paradoxais focando no caráter

contraditório e na circularidade presentes em tais argumentos.

A despeito de suas particularidades, o paradoxo de Curry cumpre o mesmo fundamental papel da ampla lista de resultados obtidos no que se convencionou chamar de *crise dos paradoxos*, a saber, revelar as entranhas da nossa ingênua e problemática compreensão de conceitos fundamentais para as ciências formais e para a racionalidade em geral. Falo aqui de conceitos tais como o de conjunto, verdade, propriedade, etc.

Referências

- Barwise, J. & Etchemendy, J. (1984) *The Liar*. New York: Oxford University Press.
- Beall, Jc (2008) “Curry’s Paradox”. In Stanford Encyclopedia of Philosophy. Online em: <http://plato.stanford.edu/entries/curry-paradox/>, último acesso: 02/03/2017.
- Boolos, George. (1998) *Logic, Logic and Logic*. Cambridge, Mass.: Harvard University Press.
- Curry, Haskell. (1942). “The inconsistency of certain formal logics”, *Journal of Symbolic Logic* 7, pp. 115-117.
- Da Costa, N & Bezieau, J-Y & Bueno, O. (1998) *Elementos de teoria paraconsistente dos conjuntos*. Coleção CLE, vol. 23. Campinas: UNICAMP.
- Glanzberg, Michael (2001). “The liar in context”, *Philosophical Studies* 103:217-251.
- Gupta, A. & Belnap, N. (1993). *The Revision Theory of Truth*. Cambridge, MA: MIT Press.
- Haack, Susan (2002) *Filosofia das Lógicas*. São Paulo: UNESP.
- Kripke, Saul (1975). “Outline of a theory of truth”, *Journal of Philosophy* 72:690-716.
- Parsons, Charles (1974). “The Liar Paradox”, *Journal of Philosophical Logic* 3: 381-412.
- Potter, Michael. (2004) *Set theory and its philosophy*. Oxford; Oxford University Press.
- Prior, Arthur N. (1955) “Curry’s Paradox and 3-Valued Logic”, *Australasian Journal of Philosophy* 33:177-82.
- Russell, Bertrand. (1902) “Letter to Frege”. in van Heijenoort (1967), pp. 124-5.
- Sainsbury, R. M. (2009) *Paradoxes*. Cambridge: Cambridge University Press.
- Tarski, Alfred. (1933) “The Concept of Truth in Formalized Languages”, in Tarski (1983).

_____. (1944) “The Semantic Conception of Truth and the Foundations of Semantics”, in *Philosophy and Phenomenological Research*, 4: 341–376.

_____. (1983) *Logic, Semantics, Metamathematics*. Indiana: Hackett Publishing Company.

vanHeijenoort, J. (ed.) (1967) *From Frege to Gödel*. Cambridge, Mass.: Harvard University Press.

Yablo, Stephen. (1993). “Paradox without self-reference”. *Analysis* 53: 251–52.

O Paradoxo Inferencial de Lewis Carroll

Lewis Carroll's Inferential Paradox

Rodrigo Reis Lastra Cid
Faculdade Arquidiocesana de Mariana

Resumo

O objetivo deste artigo é apresentar o paradoxo de Lewis Carroll sobre a justificação de princípios lógicos, tal como também algumas tentativas de solução para ele. Isso é importante, pois se há princípios lógicos básicos, parece-nos necessário haver também uma justificação para tais princípios. Levando em consideração algumas observações de Ryle, Devitt e Kripke sobre o tema, pretendemos apresentar brevemente suas teorias e suas principais críticas umas às outras e, principalmente, as críticas contra a teoria da adoção.

Palavras-Chave

Lógica. Paradoxo Inferencial. Lewis Carroll.

Abstract

My main aim at this paper is to present Lewis Carroll's Paradox on the justification of logical principles inasmuch as some attempts of solving it. This is important because if there are basic logical principles, it also seems necessary to exist some justification for them. By considering some observations from Ryle, Devitt and Kripke about the theme, we intend to briefly display their theories and their core critics among themselves and, mainly, the critics against adoption theory.

Key-Words

Logic. Inferential Paradox. Lewis Carroll.

Há um intrigante problema para as demonstrações e justificações de princípios lógicos básicos (axiomas lógicos), o qual seria justamente o problema apontado pelo paradoxo de Lewis Carroll (1895), em "What the tortoise said to Achilles". Grosso modo, o paradoxo mostra que ter os axiomas ou regras de inferência de uma lógica como premissas em um argumento não é suficiente para justificar a própria regra de inferência. A ideia do paradoxo é que se, numa certa inferência, o *modus ponens* não for assumido como uma regra de inferência para além das premissas de um argumento, não poderemos realizar a inferência. O paradoxo trata de uma conversa entre a tartaruga e Aquiles – o que nos remete aos paradoxos de Zenão, dos quais alguns, sobre a impossibilidade do movimento, eram expressos por meio de histórias entre a tartaruga e Aquiles¹ - e ele nos mostra não a impossibilidade do movimento, mas a impossibilidade de realizar uma inferência.

O caso é o seguinte: a tartaruga desafia Aquiles a obrigá-la logicamente a aceitar uma determinada inferência. A inferência avaliada no paradoxo é a de A e B para Z.

(A) coisas que são iguais ao mesmo são iguais uma à outra;

(B) os dois lados deste triângulo são duas coisas que são iguais ao mesmo;

I - Um exemplo de Zenão seria a corrida entre a tartaruga e Aquiles: Aquiles dá uma margem de distância para a tartaruga, e isso o impede, para sempre, de ultrapassar a tartaruga, pois no tempo em que Aquiles chega à metade da distância com a tartaruga, a tartaruga já terá andado um pouco, e assim que Aquiles chegar novamente à metade da nova distância, a tartaruga já terá andado mais um pouco, e assim por diante.

(Z) os dois lados deste triângulo são iguais um ao outro.

Se, ceticamente, não assumirmos a regra do *modus ponens*, podemos pensar que Z não se segue de A e B. Se assim o pensarmos, não adianta adicionarmos uma premissa C, expressando, de alguma forma, o *modus ponens*, por exemplo, expressando que Z se segue de A e B (C: se A e B, então Z). Pois daí poderíamos duvidar que, de A, B e C, segue-se Z – o que, para ser aceito, nos obrigaria a aceitar D (D: se A, B e C, então Z). Mas daí surgiria uma análoga dúvida sobre se Z se segue de A, B, C e D. Como esse é entendido como um questionamento cético por parte da tartaruga, não podemos responder, dizendo que a lógica de primeira ordem é monotônica, [i.e., $(\Gamma \vdash \varphi \wedge \Gamma \subseteq \Delta) \rightarrow (\Delta \vdash \varphi)$], pois, no ponto em que está a tartaruga (esse ponto cético), isso não pode ser pressuposto.

De um modo um pouco menos complexo, diríamos que esse é um paradoxo de uma pessoa que parece não conseguir realizar uma derivação, a não ser que aceite uma regra; mas para aceitar essa regra, acaba tendo de aceitar uma regra para aceitar essa regra, e assim por diante. Vejamos um novo exemplo, com outra versão do *modus ponens*:

- (1) Se eu gosto de gatos, vivo em casa.
- (2) Eu gosto de gatos.
- (3) Logo, vivo em casa.

A ideia é que uma pessoa que aceita 1 e 2 tem de, obviamente, aceitar 3. Caso alguém aceite 1 e 2, mas rejeite 3, ela não poderia continuar rejeitando 3, pensa Aquiles, se adicionarmos a premissa B, que seria: “se 1 e 2 são verdade, então 3 é verdade”. Teríamos assim:

- (1) Se eu gosto de gatos, vivo em casa.
- (2) Eu gosto de gatos.
- (B) Se 1 e 2, então 3.
- (3) Logo, vivo em casa.

Mas o problema colocado pela tartaruga estaria só começando, pois alguém poderia

aceitar 1, 2 e B, mas rejeitar 3. Daí teríamos de adicionar C: “se 1, 2, B, então 3”. Tendo assim:

- (1) Se eu gosto de gatos, vivo em casa.
- (2) Eu gosto de gatos.
- (B) Se 1 e 2, então 3.
- (D) Se 1, 2, B, então 3.
- (3) Logo, vivo em casa.

O Paradoxo de Carroll mostra-nos um regresso ao infinito, se decidirmos pedir por uma regra explícita de derivação que nos explique as derivações mais básicas. Supostamente, derivações como *modus ponens*, *modus tollens*, entre outros, são regras de derivações básicas, e não há regras mais básicas que façam um usuário que ainda não entendeu que 3 deriva-se de 1 e 2 entender a razão. A solução para esse paradoxo, na literatura, é transformar a lógica em meramente sintática, removendo todo seu caráter semântico. Se a aceitarmos como um sistema de regras puramente formais, não temos de ter regras para aplicar regras *ad infinitum*; só precisaríamos aceitar as regras básicas. Estas por sua vez ficariam sem explicação – o que não necessariamente é algo deveras ruim, já que os próprios axiomas lógicos fundamentais não parecem ter eles mesmos fundamento algum para além de si mesmos.

A moral da história parece ser que se estamos argumentando, não poderemos provar as regras de inferência, utilizando as próprias regras de inferência. Mas se suspendermos o nosso juízo sobre as regras básicas de inferência, não poderemos realizar mais nenhuma inferência e, conseqüentemente, não poderemos provar regra de inferência alguma, pois toda prova pressupõe um método de inferência que sirva de prova. Esse paradoxo parece ser impossível de evitar, se a lógica não for formalizada por meio de axiomas e regras de inferência como coisas distintas. Um tremendo quebra-cabeças para filósofos e lógicos.

Romina Padro (2015), em *What the tortoise Said to Kripke: the Adoption Problem and the Epistemology of Logic*, tenta mostrar que temos mais razões para crer que as nossas práticas inferenciais devem ser justificadas não por meio de proposições ou axiomas que adotamos, mas por meio de regras que não adotamos propriamente, mas utilizamos de alguma forma. Ela diria, junto com Ryle (1945), que há uma diferença entre *saber que* algo é verdadeiro e *saber como* fazer alguma coisa, e que o modo como sabemos as regras de inferência se assemelha mais a uma capacidade para inferir do que a um conhecimento proposicional

que, ao ser adotado, permitiria a inferência.

Padro (2015) nos fala de 3 problemas inter-relacionados, na epistemologia da lógica, para lidar com o paradoxo de Carroll – que são: (1) o problema da justificação: justificar princípios lógicos; (2) o problema do intitlamento: justificar o uso das inferências e a transmissão da verdade em lógica; (3) o problema da natureza da inferência: dizer o que é formar uma crença a partir da inferência sobre outras e em que condições isso é possível. Em termos gerais, perguntamo-nos qual o tipo de relação que tem de haver entre premissas e conclusão, para que a inferência seja realizada?

Kripke enfoca sua resposta no problema 3, o problema da natureza da inferência, em seu livro *Wittgenstein on Rules of Private Language* e em algumas palestras não publicadas (1974a, 1974b, 1982), a fim de refletir sobre as condições que tornam possível ou impossível alguém realizar uma inferência básica. A esse problema, ele chama de *Problema da Adoção*: se x infere de acordo com a lógica, então não é necessária a adoção de nenhum princípio adicional; se x não infere de acordo com a lógica, então não é possível a sua adoção; de todo modo, certos princípios lógicos básicos não podem ser adotados.

A interpretação comum é a de que não podemos ver as regras de inferência como premissas adicionais no argumento, que precisariam de justificação independente, pois se o fizéssemos, cairíamos num regresso ao infinito. Assim, temos de distinguir implicação lógica (*entailment*) de regras de inferência. Russell (1903, 1912) pensa também que devemos distinguir a implicação lógica da implicação material. Enquanto no segundo nada é afirmado para além da sentença condicional, no primeiro afirmamos cada uma das fórmulas que estão à esquerda e à direita do martelo sintático. Os filósofos tentam precisar essa distinção, pois é um teorema que $(\Gamma, A \vdash B) \rightarrow (\Gamma \vdash A \rightarrow B)$, e, assim, não parece tão despropositado tentar transformar uma regra de inferência em um teorema dentro de uma derivação. Peter Winch, um wittgensteiniano contemporâneo, pensa, não lembro em que obra, que o paradoxo nos mostra que o processo de inferência não pode ser completamente representado por uma fórmula lógica – o que seria um problema sério para o convencionalismo em lógica; para ele, no espírito de Ryle, aprender a inferir é um *saber como fazer* antes do que um conhecimento proposicional, e o saber como fazer não é redutível ao conhecimento proposicional, i.e., Winch rejeita tal redutibilidade, tese definidora da teoria do intelectualismo.

Costumamos aceitar que uma regra de inferência é válida sse as premissas implicam a conclusão. A pergunta da tartaruga inverte a situação, querendo saber quando “as premissas implicam a conclusão” é válido? Certamente não podemos responder a isso, dizendo “sempre que as premissas implicarem a conclusão” e nem “sempre que tivéssemos uma regra de inferência”, pois isso seria circular. Há uma série de diagnósticos diferentes para o

paradoxo, mas todos eles envolvem dizer que uma distinção importante foi negligenciada, e que, nos termos de Kripke, há uma diferença significativa entre (i) aceitar racionalmente e proposicionalmente uma crença e (ii) adotar uma crença por razões [que podem ser pragmáticas]; e, se essas razões forem a aceitação racional de um princípio lógico (ou de uma regra de inferência), então não será possível adotar a regra.

Como nos lembra Padro (2015), para provar seu ponto, Kripke nos pede para pensar em um indivíduo chamado “Harry”, que nunca realizou uma inferência que instancie o *modus ponens* ou a instanciação universal. Assim, Harry não passa de

- (1) Todos os animais do filme Madagascar falam, e
- (2) Alex é um animal do filme Madagascar,

para

- (C) Alex fala.

Se Harry não consegue derivar C a partir de 1 e 2, então de nada adiantaria adicionar a regra da instanciação universal como premissa (3), dizendo que todas as proposições universais implicam cada uma de suas instâncias. E não adianta, pois para reconhecer que a regra 3 é aplicável à proposição 1, é necessário antes já saber realizar instanciações universais, pois reconhecer que uma proposição (a proposição 1) é instância de uma proposição universal (a proposição 3) já é realizar uma instanciação universal, a saber, a instanciação do esquema que expressa a regra. Nesse caso, não há nenhuma informação que falta a Harry, que, se ele a tivesse, poderia realizar adequadamente a inferência. O que falta a ele, nos termos de Ryle, seria saber como realizar a inferência, e não ter de fato o conhecimento proposicional da regra de inferência. A diferença entre o caso de Harry e o de Aquiles é que o caso de Harry não o faz entrar em regresso, mas o faz sempre ter de pressupor a regra a fim de poder aprendê-la; além disso, ele é mais sobre um princípio geral do que sobre uma inferência particular.

O caso de Harry visa a desafiar as teorias que explicam os atos inferenciais por meio da aceitação racional de princípios e visa a mostrar que, mesmo se modificarmos a teoria de aceitação racional para a teoria de seguir regras, ainda assim, não há como ajudar Harry a

inferir – seja o que for que queiramos dizer com “seguir regras”. Kripke mesmo diria que internalizamos tais princípios lógicos e daí conseguimos produzir muitos casos particulares desse princípio. Sendo inatos ou introduzidos por indução, esses princípios formam a base do nosso conhecimento inferencial. O problema dessa virada psicologista é que traz à tona, com bastante força, o problema da justificação dos princípios lógicos, se eles forem fundados em algo inato ou induzido.

Toda teoria que tente explicar a inferência como fruto da aceitação de certos princípios lógicos acaba sendo por demais exigente, dado que a *logica utens* não se identifica com a *logica docens*: as pessoas têm uma prática inferencial a qual não sabem justificar e elas não sabem reconhecer proposicionalmente os princípios que estão aplicando.² Com relação a Harry, os intelectualistas gostariam que ele soubesse seguir a regra por tê-la aceitado, mas cada um de nós consegue dizer de si mesmo que tem essa razão? Como nos parece lembrar Wittgenstein (1953, 1964), seguir uma regra e pensar estar seguindo uma regra são duas coisas distintas, tal como agir conforme a uma regra e de fato seguir uma regra. Podemos agir conforme a uma regra por acaso. Assim, o que seria “seguir uma regra”? Certamente não podemos exigir que a pessoa saiba expressar qual regra está seguindo, pois a maioria dos leigos em lógica não reconheceria nem a fórmula do *modus ponens* e da instanciação universal, quanto mais saber expressá-las.

Ao menos quatro respostas são possíveis para o que é seguir uma regra: (I) a aceitação da regra é tácita, e é um estado mental que codifica a regra, acessível indiretamente pelo sujeito que realiza a inferência; (II) as regras estão programadas [*hardwired*] em nós e governam o processo de inferência sem serem representadas; (III) os atos inferenciais são explicados por se fundarem numa capacidade primitiva, que não é informada pelas regras inferenciais, mas que uma prática compartilhada pode aperfeiçoar a capacidade; (IV) as inferências se justificam por se fundarem em aceitação proposicional explícita (até agora, a teoria mais problemática).

O problema aqui parece ser que se a aceitação proposicional explícita (i.e., IV) não pode ajudar Harry, então parece que nada na aceitação tácita da regra (i.e., I) poderia também. Com relação a regras programadas e capacidade primitiva (i.e., respectivamente II e III), se Harry já não tiver as regras de inferência programadas em si ou não tiver essa capacidade básica, então parece que não há nada que possamos fazer que o leve a inferir – se houvesse, seria antes em termos de programar Harry de alguma forma ou fazer-lhe uma cirurgia cerebral do que em termos de aumento de conhecimento proposicional ou modificação dos teoremas aceitos. Além disso, II e III nos fazem falar que somos antes governados pelas

2 - (i) *Logica utens*: a lógica de fato usada pelas pessoas, e (ii) *Logica docens*: princípios lógicos formalizados ou estudados (Padro, 2015, p. 45-46).

regras de inferência do que realizamos um processo de seguir regras.

Um outro caso, semelhante ao de Harry, é o caso de *Earnie e a Instanciação Perversa* (1974a). O padrão de inferência de Earnie é que a partir de

$$\forall x (Fx)$$

ele deriva

$$\sim Fa, \sim Fb, \sim Fc, \dots$$

ou seja,

$$\forall x (Fx) \vdash \sim Fa, \sim Fb, \sim Fc, \dots$$

Não adiantaria dizer para ele que “*proposições universais implicam a negação de cada uma de suas instâncias*” é falacioso, pois essa é uma proposição universal e, segundo o padrão de inferência de Earnie, não será falacioso derivar que cada instância dela será falsa. Como ela também é uma instância da Instanciação Perversa, então Earnie estará intitulado a concluir que ela é falsa (i.e., que sua negação é verdadeira). No caso de Earnie também não parece haver nada que possamos fazer para que ele consiga realizar a instanciação universal corretamente. Esse caso nos confirma o problema da adoção de Kripke, de que se alguém já realiza corretamente as inferências básicas, então não há necessidade de adoção adicional de nenhum princípio e, no caso de ele não realizar, não há possibilidade de adoção de nenhum princípio.

Com relação, então, a Harry e Earnie, poderíamos tentar dizer que o que lhe falta é aprender a inferir do modo como aprendemos, por meio de exemplos. Ainda assim, teríamos os seguintes problemas:

- (i) não podemos estar certos sobre qual a regra está sendo aprendida. Wittgenstein nos lembra que não é possível aprender uma regra a partir dos

casos particulares e nem é possível determinar qual a regra que outra pessoa está seguindo;

(ii) isso não daria conta do ponto principal, que seria mostrar que um sistema axiomático, sem regras de inferência, não é capaz de formalizar toda a lógica – para dar conta, precisaríamos de um aprendizado de cima para baixo (do conhecimento proposicional para o conhecimento de como se faz algo, i.e., do *knowthat* para o *knowhow*), e não de um mero aprendizado de baixo para cima (do *knowhow* para o *knowthat*);

(iii) não garante o aprendizado e, portanto, não justifica a inferência.

Wittgenstein diria que se fosse necessária uma regra de inferência para justificar a inferência, então precisaríamos de uma regra de inferência para justificar uma regra de inferência, pois a própria regra de inferência é uma inferência – por exemplo, a forma lógica válida do *modus ponens* é apresentada como uma inferência. Além disso, temos de levar em conta que não é sempre que inferimos corretamente. Nosso raciocínio natural está cheio de equívocos e outras falácias. Kripke, por sua vez, pensa que não precisamos de regras de inferência; e que é a formalização da lógica que nos dá a errônea impressão de que precisamos. Além disso, há uma razão quineana para pensarmos que seguir uma regra não é algo que possa ser feito cegamente por qualquer um. Para entendermos uma regra de inferência, temos de entender suas partículas lógicas, que, por sua vez, só podem ser entendidas, se, antes, conhecermos lógica (*logica docens*).

Ryle, por sua vez, como é um anti-intelectualista, pensa que o *knowhow* não é redutível ao *knowthat*, e que saber fazer uma inferência é um caso de *knowhow*. Assim, enquanto Ryle pensa que

Saber como fazer φ =_{df} Ter a capacidade para φ

os intelectualistas pensam que

Saber como φ =_{df} *Saber que* x é uma forma de fazer φ

As razões de Ryle (apud Padro, 2015) contra o intelectualismo são as seguintes:

- (i) intelectualismo não dá conta da distinção entre conhecimento e sabedoria;
- (ii) fica difícil atribuir inteligência ou estupidez por um observador externo, i.e., não dá conta de explicar a diferença que parece haver entre duas pessoas com o mesmo conhecimento proposicional, mas que uma é tola e outra é esperta;
- (iii) explica atos de inteligência (*knowhow*) apelando para atos de inteligência anterior (*knowthat*) – o problema aqui é que se precisássemos de atos de inteligência para explicarmos atos de inteligência, então entraríamos num regresso vicioso;
- (iv) se um feito, para ser inteligente, precisa ser julgado pela consideração de uma proposição regulativa, então há um espaço entre a consideração da proposição e sua efetiva execução (um espaço que permite inclusive a não execução da proposição considerada), mas esse espaço não pode ser explicado pela consideração anterior de outra proposição, sem cairmos em outro círculo vicioso, e nem pelo próprio feito, sem sermos circulares.

Ryle (1945) rejeita que um conhecimento proposicional tácito ou implícito explique o ato inferencial (rejeita I e rejeita IV), por causa do *problema da aplicação*: que é explicar como *saber a regra* se relaciona com *saber como aplicar a regra* – duas coisas muito diferentes, tal como qualquer estudante inicial de lógica pode comprovar. A principal objeção intelectualista é dizer que para todo saber como fazer algo, podemos nos esforçar para traduzi-lo para conhecimento proposicional (por exemplo, explicando a prática), ou, a partir do conhecimento proposicional, podemos tentar adquirir o *knowhow* (por exemplo, o conhecimento adquirido a partir de manuais de instrução). O problema, então, seria como explicar esse esforço ou tentativa.

Devitt (2011), por sua vez, defende a teoria das regras programadas – a teoria que toma as regras de inferência como programadas em nós, tal como as regras de uma calculadora. Sua intenção é nunca multiplicar as representações conscientes para além do necessário. Por isso, ele pensa a habilidade linguística como uma forma de *knowhow*, e a habilidade de inferir como um tipo de habilidade linguística. Devitt nos lembra que os animais sabem como fazer as coisas sem saber o conhecimento proposicional físico envolvido no processo. Os principais problemas para a teoria de Devitt são: (i) a maioria de nós infere muito mal, de modo que se há uma programação lógica, ela não é muito bem feita, e (ii) não há solução para Harry e Earnie a não ser cirurgia cerebral ou algo nesses moldes. Podemos responder

a i, dizendo que inferimos mal, pois somos influenciados por conteúdo, mas o problema ii ainda estará por resolver.

Distintamente, Kripke diria que nossas inferências são baseadas nos nossos hábitos inferenciais, i.e., nossa linguagem é baseada nos nossos jogos de linguagem. Mas isso também não parece salvar Harry de seus problemas, já que se a atual exposição aos jogos de linguagem não o ajuda a inferir corretamente, não parece que mais exposição o ajudaria, e não parece haver outra forma de “adentrar” um jogo de linguagem senão jogando-o ou sendo exposto a ele.

Mas como tudo isso responde ao paradoxo de Carroll? O que parece é que qualquer resposta ao problema 3 do Paradoxo de Carroll seria uma admissão de que não há nada que fazer, por meio da linguagem, que possa obrigar logicamente a tartaruga a aceitar a inferência proposta. Se ela não consegue realizar a inferência (ou se é cética o suficiente, para não a aceitar sem justificativas independentes), algo para além de exposição aos jogos de linguagem das sociedades, para além de regras programadas, para além de capacidades primitivas e para além de aceitação (tácita ou não) de princípios básicos, deve ser apresentado como o mecanismo adequado. Na falta de um, suspendemos o nosso juízo sobre a possibilidade de o Paradoxo de Carroll ser solucionado.

O primeiro problema – o da justificativa dos princípios lógicos – é colocado em causa pelo próprio paradoxo. O segundo problema – o do intitulado do uso de princípios lógicos – parece resolvível por meio de considerações pragmáticas, embora sem solucionar o desafio *cético* da tartaruga. O terceiro problema – o de responder o que é uma inferência e em quais condições ela é possível – está sendo discutido pelas teorias indicadas aqui. Todas elas têm a dificuldade de não conseguirem explicar como podemos vencer o desafio cético da tartaruga, dizendo-nos somente em que condições uma inferência é possível (por exemplo, não é possível, se estivermos investigando ceticamente a adoção, tácita ou não, de certo princípio lógico). Assim, parece que a tarefa de justificar os nossos princípios lógicos, fazendo com que isso fundamente seu uso e sua possibilidade, ainda persiste como um quebra-cabeças a ser resolvido dentro da filosofia da lógica.

Referências

- Carroll, L. (1895). “What the Tortoise Said to Achilles”. *Mind* 4 (14): 278-280.
- Devitt, M. (2011). “Methodology and the Nature of Knowing How”. *Journal of Philosophy*

108 (4): 205-218.

_____. (2012). "Linguistic Knowledge". *Philosophical Inquiry* 36 (1): 314

Kripke, S. (1974a). Princeton Lectures on the Nature of Logic. Transcription of tapes. Manuscript.

_____. (1974b). The Question of Logic. Transcription of a lecture given at the University of Pittsburg. Manuscript

_____. (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press.

Padro, R. (2015). *What the Tortoise Said to Kripke: the Adoption Problem and the Epistemology of Logic*. Dissertations, Theses, and Capstone Projects. New York, USA, CUNY Academic Works.

Quine, W. V. O. (1966). *The Ways of Paradox and Other Essays*. RandomHouse. Edição aumentada e revisada, Harvard University Press, 1976.

_____. (1970). *Philosophy of Logic*. Harvard University Press.

Russell, B. (1903). *The Principles of Mathematics*, Vol. 1. Cambridge University Press.

_____. (1912). *The Problems of Philosophy*. Barnes & Noble Books.

Ryle, G. (1945). "Knowing How and Knowing That". *Proceedings of the Aristotelian Society* 46: 1-16.

Wittgenstein, L. (1964). *Remarks on the Foundations of Mathematics*. von Wright, G. H., Rhees R., and Anscombe G. E. M. (eds.). Blackwell.

Wittgenstein, L. (1953). *Philosophical Investigations*. Anscombe G. E. M. and Rhees R. (eds.), Blackwell.

O problema da vagueza

The problem of vagueness

Sagid Salles¹

Programa de Pós-Graduação Lógica e Metafísica da UFRJ; Universidade Federal do Amazonas.

Resumo

Este artigo é uma introdução crítica ao problema da vagueza. Não vou explicar ou avaliar as teorias da vagueza aqui. Meu objetivo é introduzir e discutir o próprio problema, com o intuito de alcançar uma formulação clara do mesmo. Minha formulação tornará claro quais condições uma teoria ideal da vagueza deveria satisfazer, assim como os principais obstáculos para desenvolver teorias que satisfaçam estas condições.

Palavras-Chave

O problema da vagueza. Paradoxo sorites. Critérios de adequação.

¹ - Agradeço a Guido Imaguire e Elizielly Martins pela leitura e sugestões quanto ao presente artigo. Aos membros do NULFA (Núcleo de Estudos em Lógica e Filosofia Analítica), da UFAM, pela paciência em discutir e criticar as ideias apresentadas aqui. Agradeço também à CAPES pelo financiamento que tornou possível a produção deste artigo.

Abstract

This paper is a critical introduction to the problem of vagueness. I will not explain or evaluate the theories of vagueness here. Rather, my goal is to introduce and discuss the problem itself, in order to propose a clear formulation of it. My formulation will make it clear which conditions an ideal theory of vagueness should satisfy, as well as the main obstacles to developing theories which satisfy these conditions.

Key-Words

Problem of vagueness. Sorites paradox. Criteria of adequacy.

Introdução

Há algum tempo conversava com um amigo sobre o problema filosófico da vagueza e comentava que as principais teorias existentes eram igualmente implausíveis. Ele retrucou-me que as principais teorias existentes eram na verdade igualmente plausíveis. Apesar de as duas afirmações serem literalmente equivalentes, tínhamos diferentes coisas em mente. Eu queria ressaltar algo negativo nas principais teorias, enquanto ele queria ressaltar algo positivo. Numa coisa podíamos concordar, dado que as teorias são muitas vezes inconsistentes entre si, havia uma flagrante dificuldade de escolher entre elas. Mesmo que nos limitemos às principais teorias a situação é bastante complicada. Entre as principais teorias da vagueza estão as teorias trivalentes, Supervaloracionismo, Teoria dos Graus de Verdade, Epistemicismo e o Contextualismo. Não há uma delas que possamos considerar *a teoria a ser refutada*; todas elas aparecem em posição similar no debate filosófico contemporâneo.

A despeito de minha impressão pessimista sobre as principais teorias existentes, eu não vislumbrava qualquer modo alternativo de resolver o problema. Uma opção seria tornar-me cético quanto à possibilidade de encontrarmos uma solução ou, quem sabe, rejeitar minhas impressões quanto ao tema. Me parecia, contudo, que o caminho mais promissor seria esquecer momentaneamente as teorias e repensar o problema. Você provavelmente já sentiu que um problema é tão difícil de resolver, e as soluções propostas tão implausíveis,

que o melhor caminho seria voltar para o início e repensar o problema. Essa é a estratégia adotada neste artigo.

Este artigo é uma introdução ao problema filosófico da vagueza. Não se trata, contudo, de uma introdução no sentido mais comum, contendo uma apresentação rápida do problema e uma exposição informal das principais soluções existentes. Ao contrário, meu foco é exclusivamente *o problema da vagueza* e não as suas soluções. Segundo, trata-se de uma introdução crítica, que vai além de meramente apresentar o modo como o problema da vagueza é geralmente entendido. Ao contrário, o objetivo principal do artigo é apresentar uma formulação mais clara do problema, que sirva como uma base segura para começarmos a propor soluções. A formulação proposta clarificará quais condições devemos esperar que uma teoria da vagueza satisfaça; além disso, a discussão revelará as principais dificuldades para a satisfação das mesmas. Desse modo, o resultado também será interessante para aqueles cujo objetivo é avaliar ou construir teorias da vagueza.

A formulação final do problema aparecerá apenas na última seção do artigo, após uma longa discussão sobre vários aspectos do problema. O artigo é dividido em quatro seções. Na primeira, apresento o paradoxo sorites em uma versão informal, explico a sua importância e extensão e introduzo algumas restrições iniciais quanto ao modo como entenderei o problema da vagueza aqui. Na segunda, apresento um pouco da história do paradoxo sorites, e três formulações mais precisas do mesmo. Sustentarei que, tão logo começamos a discutir sobre o paradoxo e as alternativas de soluções, percebemos o surgimento de três critérios de adequação para uma teoria ideal da vagueza. Um dos critérios, o chamado *critério da precisificação*, é o mais controverso e será discutido em algum detalhe. Na terceira seção, apresento um conjunto de intuições normalmente associadas ao fenômeno da vagueza, intuições estas que toda teoria deve ser capaz de sistematizar. Por fim, apresento a formulação do problema, que acredito surgir naturalmente – e espero que o leitor perceba o mesmo – da discussão das três seções anteriores.

1. Preliminares

Cotidianamente, a palavra “vagueza” é usada de diferentes modos. Tomando um exemplo de Eklund (2005, p.27), suponha que você me pergunte onde João está e eu lhe responda que ele está em algum lugar entre Manaus e Rio de Janeiro. Nesse caso, você pode dizer que minha resposta é vaga, no sentido de ser pouco específica sobre a localização de João. Outras vezes “vagueza” é usada como significando *incerto*, como quando se diz que

o *nosso destino é vago* ou que o réu deve ser absolvido, devido à vagueza das imputações contra si. O uso comum da palavra “vagueza” diz respeito a uma variedade de fenômenos. Em contrapartida, os filósofos pretendem falar de vagueza com um fenômeno específico em mente. O que mais chama a atenção sobre este fenômeno é que ele está por trás do que é conhecido como *paradoxo sorites*. Passemos a uma exposição informal do paradoxo.

1.1. Sorites (formulação informal), sua extensão e sua importância

Imagine uma pessoa qualquer, chame-lhe “João”, e suponha que os cabelos dela nasçam de um em um. Em um primeiro momento, João terá 0 fios, em seguida terá 1, depois 2, 3, 4, etc. Considere a seguinte sequência abaixo, com cada linha descrevendo um diferente momento de João, representando cada novo fio de cabelo.

João tem 0 fios de cabelo na cabeça

João tem 1 fio de cabelo na cabeça

João tem 2 fios de cabelo na cabeça

.

.

.

João tem 5.000 fios de cabelo na cabeça

João tem 5.001 fios de cabelo na cabeça

.

.

.

João tem 10.000 fios de cabelo na cabeça

Pare na primeira linha da sequência acima, que descreve João com 0 fios de cabelo na cabeça. Pergunte-se: João é careca? A resposta, óbvia, é que “sim”. Passe então para a segunda linha, e faça novamente a pergunta: João é careca? A resposta, novamente, é que

“sim”. Vá para a terceira e repita o procedimento. Novamente terá “sim” como resposta. Imagine-se repetindo o procedimento indefinidamente; para cada linha terá de responder se João é ou não careca. Pergunte-se: quando a resposta deixará de ser “sim” e começará a ser “não”? Intuitivamente, a adição de um mero fio de cabelo na cabeça de um careca não é suficiente para torná-lo não-careca. Se isto é assim, então você nunca poderá parar de responder “sim” e, conseqüentemente, será forçado a considerar que João é careca mesmo quando tem 10.000 fios de cabelo em sua cabeça. Este resultado é difícil de engolir.

Repita o procedimento de baixo para cima. Comece considerando a linha 10.000 e pergunte-se: João é careca? A resposta intuitiva é “não”. Faça a mesma pergunta na linha 9.999. Novamente, a resposta será “não”. Passe linha por linha repetindo o procedimento. Quando é que a resposta deixará de ser “não” e passará a ser “sim”? Intuitivamente, a subtração de um mero fio de cabelo não é suficiente para fazer com que um não-careca torne-se careca. Conseqüentemente, você nunca poderá mudar a sua resposta, de modo que será forçado a considerar João não-careca mesmo quando tem 0 fios de cabelo na cabeça.

No primeiro caso, começamos com a suposição inocente de que uma pessoa com 0 fios de cabelo na cabeça é careca, e terminamos com a conclusão surpreendente de que todos os outros na sequência – até mesmo uma pessoa com 10.000 fios – também são carecas. No segundo caso, começamos com a suposição inocente de que uma pessoa com 10.000 fios é não-careca, e terminamos com a conclusão de que todos os outros na sequência – mesmo aqueles com 0 fios de cabelo na cabeça – também são não-carecas. O tipo de argumento usado para passar das referidas suposições inocentes para as conclusões surpreendentes é conhecido como *paradoxo sorites*. Tomados em conjunto, os dois argumentos acima nos permitem mostrar que todo mundo é careca e não-careca ao mesmo tempo.

Qual é a importância disso? Há alguns anos, quando estava na graduação em filosofia, fui a uma comunicação sobre o paradoxo sorites. Um dos ouvintes insistia que não se tratava de um problema filosófico. Uma de suas razões foi que seria simplesmente absurdo, por exemplo, um filósofo doutorado em *calvície*. Mais recentemente, eu próprio apresentava uma palestra sobre o tema e dois ouvintes passavam quase todo tempo a se olhar e rir. Mais tarde descobri a razão: parecia-lhes simplesmente inacreditável alguém perder tempo pesquisando algum problema filosófico relacionado à *carequice*. No primeiro caso, duvidou-se de que se tratava de um problema filosófico, no segundo duvidou-se de que se tratava de um problema importante.

A preocupação com o paradoxo sorites não tem nada a ver com a preocupação com o problema da calvície. O que chama a atenção dos filósofos nos argumentos acima é que se trata de um *paradoxo*, tanto no sentido fraco quanto no sentido forte do termo. No sentido

fraco, um paradoxo é um argumento aparentemente válido, com premissas aparentemente verdadeiras e conclusão aparentemente falsa. Se considerarmos cada versão isolada do sorites acima, temos um paradoxo no sentido fraco. No sentido forte, um paradoxo é um argumento aparentemente válido, com premissas aparentemente verdadeiras e conclusão contraditória. Se considerarmos conjuntamente ambas as direções do sorites acima, temos um paradoxo no sentido forte. Paradoxos nos indicam que alguma coisa pode estar errada ou com nossas regras de inferência ou com crenças que muitas vezes consideramos intuitivas. Filósofos estão interessados em resolver o paradoxo sorites, encontrando onde está o erro, se é que há algum erro.

Uma das coisas que impressiona neste paradoxo é o quão extensa é a sua aplicação. Considere noções como *magro, gordo, alto, baixo, grande, pequeno, maduro, criança, adulto, vivo, morto, forte, fraco*, etc. Para cada uma delas, você pode formular uma versão do argumento acima. Eis um exemplo com a noção de *alto*. Seja João um ser humano adulto, e imagine a seguinte sequência abaixo.

João tem 1,50 m de altura.

João tem 1,51 m de altura.

João tem 1,52 m de altura.

.

.

.

João tem 2,00 m de altura.

Pare na primeira linha da sequência e pergunte-se: João é alto? A resposta é “não”. Repita o procedimento na segunda linha, terceira, quarta, e assim por diante, passando por cada linha e fazendo-se a mesma pergunta. Quando é que a resposta deixará de ser “não” e passará a ser “sim”? Ora, não parece que um mero centímetro é suficiente para fazer com que alguém não-alto torne-se alto. Resultado: você nunca poderá mudar sua resposta, sendo forçado a considerar que João é não-alto mesmo quando tem 2,00 metros de altura. Por outras palavras, todos na sequência serão não-altos. Repita o procedimento de baixo para cima. O resultado será que todos na sequência são altos. Com as duas versões combinadas, temos que todos são altos e não-altos ao mesmo tempo.

Considere a sequência.

João tem 0 anos de idade.

João tem 0 anos + 1 segundo de idade.

João tem 0 anos + 2 segundos de idade.

João tem 0 anos + 3 segundos de idade.

.

.

.

João tem 0 anos + 3.153.600.000 segundos de idade (aproximadamente 100 anos de idade)

Com argumentos similares aos já apresentados, pode-se mostrar que todo mundo na sequência acima é criança e não-criança ao mesmo tempo, que todos são maduros e não-maduros ao mesmo tempo, etc. Em suma, para qualquer das noções acima, podemos formular um argumento semelhante que nos leva a uma conclusão contraditória. Na verdade, é bastante plausível que a maior parte das noções de que dispomos é suscetível ao paradoxo sorites. Não se trata, portanto, de um paradoxo acerca de uma noção específica. Falhar em entender isto é não apreciar corretamente o problema.

De fato, talvez o problema seja ainda mais radical do que parece. Primeiro porque, com um pouco de criatividade, conseguimos formular versões do argumento sorites mesmo para noções que não são obviamente vagas. Graham Priest (2004, p.9) certa vez formulou uma versão para mostrar que você é um ovo mexido, enquanto Peter Unger (1979) aplicou os argumentos sorites a noções como *pedras*, *mesas* e objetos cotidianos em geral.

Em segundo lugar, o sorites nos permite argumentos a favor de teses filosóficas que consideraríamos muito contraintuitivas. Eis um exemplo para a filosofia moral. Se o sorites está correto, podemos mostrar que qualquer pessoa, em qualquer idade, é uma criança. Ora, é imoral punir crianças legalmente; logo, é imoral punir qualquer pessoa legalmente.

Não é necessário entrar em maiores detalhes aqui. Basta notar que a vagueza é um fenômeno ao mesmo tempo muito amplo e com conseqüências muito sérias. Infelizmente, é também um problema muito difícil.

1.2. O problema da vagueza é um problema sobre o quê?

Inicialmente, pode-se dizer que explicar o fenômeno da vagueza é responder adequadamente à seguinte pergunta: qual é a natureza da vagueza? Vimos que uma resposta adequada a essa pergunta deve, no mínimo, permitir-nos resolver o paradoxo sorites. Mas isso não é suficientemente claro. Repare, por exemplo, que minha formulação sequer deixa claro se se trata de um problema sobre a linguagem ou o mundo. Na seção anterior, falei de *noções* vagas como, por exemplo, a *noção de careca*. Ao fazer isso, não especifiquei se falava da propriedade de *ser careca* ou do predicado “careca”. Não é óbvio se o problema da vagueza é sobre objetos e propriedades ou sobre a linguagem. Em todo caso, a partir de agora adoto a segunda opção. Portanto, o problema da vagueza será entendido como um problema sobre a natureza das expressões vagas. De uma forma mais precisa, até agora temos que: *o problema da vagueza é o problema de explicar as expressões vagas de uma forma que resolva o paradoxo sorites*. O leitor deve se lembrar, contudo, que é sempre possível uma formulação mais geral em termos de *noções* ao invés de *expressões*.

Em todo caso, é importante evitar confusões. A formulação do problema em termos de *expressões* não exclui, de antemão, teorias alternativas da vagueza; nem mesmo aquelas que não vêem vagueza como um fenômeno lingüístico. Considere alguém que defenda uma concepção metafísica de vagueza, isto é, que vagueza seja em última instância uma característica de objetos e propriedades do mundo. Ao se deparar com a pergunta sobre qual a natureza das expressões vagas, tem-se duas opções. (i) Pode-se dizer que o que faz com que expressões vagas sejam vagas são os objetos e propriedades vagos. Um predicado vago, por exemplo, seria vago em função de expressar uma propriedade vaga. (ii) Pode-se dizer que não há uma natureza das expressões vagas, pois não há expressões vagas de todo em todo; o que explica a nossa falsa impressão de que certas expressões são vagas são os objetos e propriedades vagas que estão relacionados com essas expressões. Em suma, mesmo aqueles que defendem que vagueza é em última instância um fenômeno metafísico terão de ou explicar a vagueza na linguagem em termos da vagueza metafísica ou rejeitar a vagueza na linguagem, explicando-a como um tipo de ilusão. Em todo caso teremos uma resposta a nossa pergunta acerca da natureza das expressões vagas. Observações similares valem para qualquer um que defenda que vagueza não é em última instância um fenômeno lingüístico.

Em suma, minha pergunta pela natureza das expressões vagas não visa excluir de antemão a possibilidade de não haver expressões vagas ou de vagueza não ser em última instância uma propriedade de expressões. O ponto é apenas tornar claro qual é o objeto de estudo daqui por diante: a linguagem.

1.3. Restrições quanto ao problema na linguagem

Focar-se na vagueza de expressões, isto é, no problema de explicar a natureza das expressões vagas, nos permite restringir a investigação. Entretanto, mesmo com esta restrição o problema fica amplo demais. O problema é que o fenômeno da vagueza não está restrito a um ou outro tipo de expressões de nossa linguagem. De fato, a vagueza abrange uma classe muito ampla de tipos de expressões. Já vimos o caso de termos gerais como “careca”, “alto” e “criança”, mas podemos ir além disso. Também termos singulares e quantificadores, por exemplo, podem ser vagos.

Considere primeiramente o caso dos termos singulares, isto é, termos usados para fazer referência a um e apenas um objeto do mundo. O nome “Monte Everest” é aparentemente vago, pois é suscetível ao paradoxo sorites. Ao que o nome “Monte Everest” se refere? Imagine uma linha desenhada em torno da montanha. Se o nome se refere ao objeto dentro dessa linha, também se refere ao objeto que está dentro de uma segunda linha que seja apenas infimamente mais larga do que a primeira, e que contorna a primeira. Repita o procedimento imaginando sucessivas linhas umas em torno das outras, cada uma apenas infimamente mais larga que a anterior. O resultado será que o nome se refere a todo o planeta Terra.

O indexical “aqui” também é aparentemente vago (ou pelo menos usado em alguns contextos de forma vaga). Considere um proferimento de (1).

(1) Aqui está chovendo.

Pergunte-se: ao que “aqui” se refere? Em alguns casos, nossos proferimentos do tipo de (1) são realizados com alguma descrição (alegadamente) precisa que fixa o referente de “aqui”. Este seria o caso se proferíssemos (1) com a intenção de nos referirmos a uma cidade ou bairro específico. Mas imagine que (1) tenha sido proferida sem alguma descrição dessas em mente, como parece de fato frequentemente ocorrer. Imagine uma distancia de 1m^2 em torno do falante. É intuitivo que “aqui” se refere àquela distância. Mas se “aqui” se refere à distância de 1m^2 em torno do falante, também se refere à distância de $1,001\text{m}^2$ em torno do falante. Repetindo o procedimento, chegamos à conclusão de que o “aqui” em (1) se refere a todo o planeta. Há ainda outros termos singulares candidatos a serem vagos, como por exemplo os

termos “ali” e “lá”. Para os nossos propósitos, contudo, os exemplos acima bastam.

Quantificadores como “muito”, “pouco”, “frequentemente”, “raramente”, etc. estão entre os casos mais intuitivos de expressões vagas. Um centavo não é muito dinheiro. Mas se um centavo não é muito dinheiro, então dois centavos também não é muito dinheiro, e assim por diante. A conclusão é que um bilhão de reais não é muito dinheiro. Uma péssima notícia para você, caso tenha um bilhão e esteja lendo isso. Versões similares de sorites podem ser facilmente formuladas para os outros quantificadores acima.

Em conclusão, não apenas termos gerais ou predicados, mas também termos singulares e quantificadores podem ser vagos. Acontece que esses são tipos muito diferentes de expressões, cada um carregando suas próprias idiossincrasias e colocando problemas particulares para uma explicação da vagueza. Compare com o que ocorre na em teoria da referência singular. Teorias da referência singular tentam explicar o fenômeno da referência dos termos singulares. Apesar de tais teorias terem seu escopo restrito aos termos singulares, existe alguma dificuldade em se apresentar uma teoria que valha para todos eles, devido ao fato de diferentes tipos de termos singulares envolverem diferentes peculiaridades. Por exemplo, enquanto a determinação do referente de termos como “aqui” e “eu” depende de alguma regra que determina um referente para cada contexto de uso (o referente de “eu”, por exemplo, será o proferidor de “eu” no contexto), a determinação do referente de nomes próprios não depende de tais regras. Uma teoria da referência para “aqui” e “eu” conterà elementos que uma teoria da referência dos nomes próprios não conterà. Peculiaridades desse tipo são obstáculos à criação de uma teoria da referência que dê conta de todos os termos singulares.

Assim como as peculiaridades de diferentes tipos de termos singulares colocam dificuldades para uma teoria geral dos termos singulares, as peculiaridades de diferentes tipos de expressões colocam dificuldades para uma teoria geral das expressões vagas. Na verdade, o caso da vagueza é ainda pior, pois envolve expressões ainda mais radicalmente distintas. Como vimos, não apenas diferentes tipos de termos singulares podem ser vagos, mas também predicados e quantificadores. A vagueza está espalhada por diferentes categorias lógicas de expressões. Pode-se esperar muita dificuldade para a construção de uma teoria geral das expressões vagas.

Por razões de simplificação, daqui por diante concentro-me nos predicados vagos.

2. O Paradoxo Sorites

2.1. Um pouco de história

Se acreditarmos nas informações de Diôgenes Laértios (D.L. II, 108), Ebulides de Mileto (Sec. IV a.c.) realmente merece o título de *mestre dos paradoxos* (Burnyeat, 1982, p.315). Ele é o descobridor de vários paradoxos além do sorites, sendo o mais famoso o paradoxo do mentiroso. Apesar de não ser unânime, a autoria de Ebulides com relação ao sorites é geralmente aceita. A expressão “sorites” vem do grego “soros”, que significa *monte* ou *pilha*, e tem relação com o fato de o paradoxo ter sido primeiramente formulado para a noção de *monte* ou *pilha*. Há dúvidas sobre se Ebulides compreendeu a generalidade e as consequências do paradoxo formulado por ele. De fato, “sorites” pode ter sido primeiramente usada como um nome para um argumento específico acerca da noção de monte ou pilha. Não é claro quando o paradoxo sorites passou a ser visto como um tipo de argumento aplicável à maior parte de nossas expressões e capaz de colocar em xeque alguns alegados princípios e verdades lógicas.

Em todo caso, há algumas coisas que podemos afirmar com segurança. Primeiro, o paradoxo tornou-se conhecido na antiguidade. Os estóicos deram considerável atenção ao sorites, sendo que Crísipos, por exemplo, escreveu pelo menos três trabalhos sobre o sorites aplicado às palavras e dois sobre *os argumentos de pouco em pouco* (D.L. VII, p.192, VII, 197), que era outro nome dado aos argumentos sorites. O sorites é também discutido por Sexto Empírico (*Outlines of Pyrrhonism*, II, 253) e Cláudio Galeno (*On Medical Experience* XVI-XVII), o último sendo talvez a discussão mais notável que chegou até nós. Em segundo lugar, não foi preciso muito tempo para que se soubesse que o paradoxo pode ser formulado para muitas outras noções além de *monte* ou *pilha*. O paradoxo apareceu em um manual estóico de lógica formulado para a noção de *pouco*, e o próprio Ebulides o formulou para a noção de *careca*. Galeno não deixa dúvidas de que estava ciente do fato de o sorites se aplicar a uma ampla variedade de noções, e não trata isso como algo novo. Havia uma compreensão razoável de que o sorites não era um problema sobre uma ou outra noção particular, mas um problema muito amplo.

As primeiras formulações do sorites, contudo, são diferentes das formulações atualmente mais comuns. Ao que parece, o sorites não foi originalmente formulado como um paradoxo (Hyde, 2011), mas como um enigma formado por uma coleção de perguntas. *Uma pessoa com 0 fios de cabelo na cabeça é careca? Sim. Uma pessoa com 1 fio de cabelo na cabeça é careca? Sim. Uma pessoa com 2 fios é careca? Sim. Quando é que a resposta vai deixar de ser “sim” e passar a ser “não”? Repare que isso não é um paradoxo, não se trata de um*

argumento aparentemente válido, com premissas aparentemente verdadeiras e conclusão aparentemente falsa ou contraditória. Na verdade, sequer se trata de um argumento.

Mas é um enigma muito difícil. Temos a intuição de que não podemos simplesmente parar de responder “sim” em algum momento e começar a responder “não”. Fazer isso nos comprometeria com a afirmação de que um mero fio de cabelo é suficiente para tornar um careca em não careca. Mas isso é um absurdo. Por outro lado, se continuarmos a responder sempre “sim” seremos comprometidos com a tese de que uma pessoa com 10.000 fios de cabelo ainda é careca e isso também é um absurdo. Assim ficamos sem saber como proceder. Um exemplo muito claro desta formulação é Galeno (*On Medical Experience* XVII, 115-116).

Nem sempre o sorites apareceu desta forma, pois apareceu como um argumento propriamente dito em um manual estóico de lógica (D.L. VII, p.82). Mignucci (1993) argumenta de forma plausível a favor de que os estóicos já dispunham de uma formulação precisa do sorites como um paradoxo propriamente dito e que a mesma estava pelo menos implícita em Galeno. Detalhes à parte, o enigma acima claramente está na origem do que hoje conhecemos por paradoxo sorites, e não é necessário muito para transformá-lo em um argumento propriamente dito. Na introdução, por exemplo, começamos com enigmas desse tipo para formular versões informais do paradoxo.

Seja como for, atualmente dispomos de formulações precisas do paradoxo sorites, nas quais ele é apresentado como um paradoxo propriamente dito. No entanto, existem algumas formulações diferentes. Isso é importante porque queremos resolver todas as versões e pode ser o caso de diferentes versões requererem diferentes soluções. Diferentes versões de um paradoxo podem envolver alguma premissa, regra de inferência ou até mesmo conclusão diferente. Suponha que resolvemos uma versão qualquer rejeitando uma regra de inferência usada na derivação da conclusão. Ora se existe uma versão diferente que não requer o uso dessa regra de inferência, então é necessário buscar por soluções alternativas para essa versão. O melhor seria uma teoria que fornecesse uma solução uniforme para todas as versões, mas a princípio não temos garantias de que isso seja sequer possível.

Nas próximas seções apresento três versões diferentes do paradoxo. As duas primeiras são as versões padrão e receberão maior atenção.

2.2. Sorites (versão condicional)

Uma pessoa com 0 fios de cabelo é careca. Se uma pessoa com 0 fios de cabelo é careca,

então uma pessoa com 1 fio também é careca. Portanto, uma pessoa com 1 fio de cabelo na cabeça é careca. Mas se uma pessoa com 1 fio de cabelo na cabeça é careca, então uma pessoa com 2 fios também é careca. Portanto, uma pessoa com 2 fios de cabelo na cabeça é careca... Repetindo o processo 10.000 vezes, chegamos à conclusão de que uma pessoa com 10.000 fios de cabelo na cabeça é careca.

O argumento pode ser dado na direção inversa. Uma pessoa com 10.000 fios de cabelo na cabeça não é careca. Se uma pessoa com 10.000 fios de cabelo não é careca, uma pessoa com 9.999 fios de cabelo também não é. Portanto, uma pessoa com 9.999 fios de cabelo na cabeça não é careca. Se uma pessoa com 9.999 fios de cabelos na cabeça não é careca, então uma pessoa com 9.998 fios de cabelo na cabeça também não é. Portanto, uma pessoa com 9.998 fios de cabelo na cabeça não é careca. Repetindo-se o processo 10.000 vezes, chegamos à conclusão de que uma pessoa com 0 fios também não é careca.

Os argumentos acima são exemplos da versão condicional do argumento sorites. Seja C_n “uma pessoa com n fios de cabelo na cabeça é careca”. O argumento, em ambas as direções acima, pode ser formulado como segue.

(SC)	(\neg SC)
C_0	$\neg C_{10.000}$
$C_0 \rightarrow C_1$	$\neg C_{10.000} \rightarrow \neg C_{9.999}$
C_1	$\neg C_{9.999}$
$C_1 \rightarrow C_2$	$\neg C_{9.999} \rightarrow \neg C_{9.998}$
C_2	$\neg C_{9.998}$
(...)	(...)
$C_{9.999} \rightarrow C_{10.000}$	$\neg C_1 \rightarrow \neg C_0$
$C_{10.000}$	$\neg C_0$

Cada caso começa com uma afirmação inocente que pouca gente disputaria, C_0 no primeiro e $\neg C_{10.000}$ no segundo. Seguindo o costume, podemos chamar a esta primeira

premissa de a *premissa categórica* de cada caso. Cada um dos casos também contém um conjunto de premissas condicionais às quais, por razões óbvias, são chamadas de *premissas condicionais*. As outras premissas são obtidas por Modus Ponens e sigo Mignucci (1993, p.236) em chamá-las de *premissas intermediárias*. A versão condicional do sorites é formada por uma premissa categórica e várias condicionais, e a conclusão é obtida por meio de *Modus Ponens*.

Nesta formulação, o sorites é um argumento propriamente dito e não um *mero* enigma. Se considerarmos cada coluna isoladamente, teremos um paradoxo no sentido de um argumento aparentemente válido com premissas aparentemente verdadeiras e conclusão aparentemente falsa. Se considerarmos as duas colunas conjuntamente, o resultado será uma infinidade de contradições. De fato, para qualquer pessoa, pode-se usar o sorites para mostrar que ela é careca e para mostrar que ela não é careca. Resultado: o predicado “careca” é incoerente.

Como não há dúvidas sobre como as premissas intermediárias são obtidas, é bastante útil apresentar uma versão resumida do argumento, na qual elas estão suprimidas. Ei-la.

(SC)	(¬SC)
C0	¬C10.000
C0→C1	¬C10.000→¬C9.999
C1→C2	¬C9.999→¬C9.998
C2→C3	¬C9.998→¬C9.997
(...)	(...)
C10.000	¬C0

Na versão resumida, resta-nos a premissa categórica, as premissas condicionais e a conclusão. Em cada caso a premissa categórica é uma suposição pouco problemática, enquanto a conclusão se segue logicamente das premissas. Mas de onde vêm as premissas condicionais? Quais razões justificam introduzir aquelas várias condicionais no argumento?

Podemos ver claramente a razão para isso quando tentamos negar alguma delas.

Suponha, por exemplo, que a condicional $C5.000 \rightarrow C5.001$ do argumento (SC) seja falsa. Isto implica que $C5.000$ é verdadeira enquanto $C5.001$ é falsa: $C5.000 \wedge \neg C5.001$. Por outras palavras, uma pessoa com 5.000 fios de cabelo na cabeça é careca e uma pessoa com 5.001 fios de cabelo na cabeça não é careca. Este resultado nos parece absurdo. Não parece que o acréscimo de um único fio de cabelo é capaz de tornar um careca em não-careca. De fato, a ideia de que uma pessoa passaria de careca para não-careca por causa do acréscimo de um único fio de cabelo é muito contra-intuitiva.

Repare que o problema aqui não é devido à condicional específica que escolhemos negar. Qualquer condicional do argumento (SC) que neguemos nos levará ao mesmo absurdo. É nisto que reside a justificativa para as condicionais: somos forçados a aceitar a verdade de cada uma delas sob pena de cair no absurdo de que o acréscimo de um mero fio de cabelo tem o poder de tornar um careca em não-careca. O mesmo vale para $(\neg SC)$. A negação de qualquer uma das condicionais em $(\neg SC)$ implicaria que a retirada de um mero fio de cabelo tem o poder de tornar um não-careca em um careca.

Crispin Wright (1975, p.334) expressou isso dizendo que predicados como “careca” são *tolerantes*. Um predicado é tolerante quando mudanças muitíssimo pequenas nos padrões relevantes de aplicação não fazem diferença para a aplicação do predicado. Se o predicado se aplicava antes da mudança, continuará se aplicando depois da mudança; se ele não se aplicava antes da mudança, continuará não se aplicando depois da mudança. O padrão relevante (ou pelo menos um deles) para a aplicação de “careca” é o número de fios de cabelo na cabeça. O acréscimo de um mero fio de cabelo na cabeça não pode tornar um careca em não-careca; se um sujeito é careca, continuará sendo careca após o acréscimo. Do mesmo modo, a retirada de um mero fio de cabelo não pode tornar um não-careca em careca; se um sujeito é não-careca antes da retirada, continuará sendo não-careca depois da retirada. O acréscimo/retirada de um único fio de cabelo simplesmente não faz diferença para a aplicação de “careca”; trata-se de um predicado tolerante.

Podemos expressar a tolerância de “careca” com o seguinte princípio.

$$(PTC): \forall n (C_n \rightarrow C_{n+1})$$

Isto significa que, para qualquer número natural n , se uma pessoa com n fios de cabelo na cabeça é careca, então uma pessoa com $n+1$ fios de cabelo na cabeça também é careca. Este princípio justifica as condicionais nos argumentos sorites acima. Repare que cada condicional em (SC) é uma instância deste princípio, enquanto cada condicional em $(\neg SC)$

é uma instância do logicamente equivalente $\forall n (\neg C_{n+1} \rightarrow \neg C_n)$.

Por fim, lembre-se que o paradoxo sorites não é restrito à noção de *careca*. A versão condicional pode ser formulada para uma infinidade de outras noções, como *monte*, *gordo*, *magro*, *alto*, *baixo*, *muito*, *pouco*, *gritar*, etc. Para cada uma destas noções, podemos formular o argumento sorites condicional mostrando que são incoerentes.

2.3. Sorites (versão indutiva)

Uma vez que estamos cientes do princípio de tolerância, podemos formular uma versão muito mais simples e direta do paradoxo sorites para o predicado “careca”. Aqui estão duas instâncias desta versão.

(SC)	(¬SC)
(1) C0	(1) ¬ C10.000
(2) $\forall n (C_n \rightarrow C_{n+1})$	(2) $\forall n (\neg C_{n+1} \rightarrow \neg C_n)$
_____	_____
(3) C10.000	(3) ¬C0

De modo menos formal, o que a versão acima de (SC) diz é o seguinte.

(SC)

- (1) Uma pessoa com 0 fios de cabelo na cabeça é careca.
- (2) Para qualquer número natural n , se uma pessoa com n fios de cabelo na cabeça é careca, então uma pessoa com $n + 1$ fios de cabelo na cabeça também é careca.
- (3) Logo, uma pessoa com 10.000 fios de cabelo na cabeça é careca.

A premissa (1) em cada caso é a premissa categórica. Desta vez, contudo, não introduzimos uma sequência de premissas condicionais, mas apenas (2), que é o princípio de tolerância. A conclusão se segue de (1) e (2) por indução matemática (por isso o nome *versão indutiva*). Novamente, temos no mínimo um argumento aparentemente válido, com premissas aparentemente verdadeiras e conclusão aparentemente falsa.

Mais uma vez, não se trata de um problema específico com o predicado “careca”, mas algo que afeta uma gigantesca parcela dos predicados de nossa linguagem. Eis um exemplo para “monte”.

(\neg SMonte)

(1) Um grão de trigo não é um monte.

(2) Para todo n , se n grãos de trigo não são um monte, $n+1$ grãos de trigo não são um monte.

(3) Logo, 10.000 grãos de trigo não são um monte.

Esta versão do paradoxo sorites tem algumas vantagens sobre a anterior. Além de ser bem mais compacta, ela torna explícita a importância do princípio de tolerância para o paradoxo. Lembre, por exemplo, que na versão anterior o princípio de tolerância estava apenas implícito, não sendo diretamente formulado no argumento. Nesta versão, contudo, ele é uma das premissas do argumento. Por essas razões, tomarei a versão indutiva como padrão, mencionando diretamente outras versões apenas quando isto tiver relevância para a discussão.

2.4. Sorites (versão da linha desenhada)

Adotando a terminologia de Hyde (2011), podemos chamar a próxima versão do sorites de a *versão da linha desenhada*. De um modo direto, pode-se apresentá-la como segue.

(1) C0

(2) \neg C10.000

$\exists n (Cn \wedge \neg C_{n+1})$

A ideia do argumento sorites acima é bastante simples. A primeira premissa afirma que uma pessoa com 0 fios de cabelo na cabeça é careca. A segunda, que uma pessoa com 10.000 fios de cabelo na cabeça não é careca. Ambas são aparentemente verdadeiras, mas disso parece se seguir que há algum número natural n que é tal que uma pessoa com n fios de cabelo na cabeça é careca e uma pessoa com $n+1$ fios não é. Esta conclusão é claramente falsa.

O princípio de tolerância não aparece nas premissas desse argumento, nem mesmo de forma implícita. Não precisamos assumir que o predicado “careca” é tolerante para aceitar que $C0$ e $\neg C10.000$ implicam que existe uma fronteira precisa entre carecas e não carecas. Imagine uma sequência de pessoas (ou uma sequência do mesmo indivíduo em diferentes momentos) indo de um sujeito com 0 fios de cabelo até um com 10.000 fios. Dado que o primeiro é claramente careca e o último claramente não-careca, tem de haver algum par de itens adjacentes na sequência no qual a transição entre carecas e não-carecas é feita; isto é, um par tal que o primeiro indivíduo é careca e o seu vizinho é não-careca.

A conclusão do argumento é logicamente equivalente à negação do princípio de tolerância, de modo que é argumentável que seja em última instância por isto que ela é absurda. Mas nem todos concordam com isto. Pode-se pensar, como Wright (2009, p.530) posteriormente fez, que a afirmação de que predicados vagos não possuem uma fronteira é uma intuição mais fundamental sobre a vagueza do que a de que predicados vagos são tolerantes. Tomando o exemplo de “careca”, a intuição fundamental sobre a sua vagueza seria que $\neg\exists n (Cn \wedge \neg Cn+1)$ e não que $\forall n (Cn \rightarrow Cn+1)$. Retornarei à discussão sobre as intuições envolvidas na vagueza no final deste artigo.

Em suma, vimos uma versão do sorites no qual ele é formulado apenas como um enigma e três nas quais se trata de um paradoxo propriamente dito. Embora essas sejam as versões mais comuns, nem de longe são as únicas. Apesar disso, meu foco será nas que foram até agora apresentadas e, mais especificamente, na versão indutiva. Nenhuma outra versão receberá um tratamento detalhado aqui. Por um lado, levar em conta todas as versões já apresentadas tornaria cada passo da discussão muito longo. Por outro lado, algumas das formulações do sorites têm motivações muito específicas que envolvem complicações que estão além de meus objetivos aqui. Michael Dummett (1975) e Wright (1975 e 1976) motivaram uma discussão sobre vagueza com interesse especial em predicados *observacionais* e na noção de *indiscriminabilidade*. Esta discussão pode envolver formulações do paradoxo voltadas especificamente para este tipo de predicados e as noções relevantes para a discussão. Uma boa apresentação resumida das diferentes versões existentes, assim como as motivações para as mesmas, pode ser encontrada em Hyde (2011).

2.5. Modos de solucionar o paradoxo

Um paradoxo é um argumento aparentemente válido, com premissas aparentemente verdadeiras e conclusão aparentemente falsa ou contraditória. Há três modos de solucionar um paradoxo.

- (a) Rejeitar uma das premissas
- (b) Rejeitar uma das regras de inferência usadas para derivar a conclusão.
- (c) Aceitar a conclusão.

Com o paradoxo sorites não é diferente. Para lembrar, eis uma instância do sorites para o predicado “careca”.

- (SC)
- (1) C0
 - (2) $\forall n (Cn \rightarrow C_{n+1})$
-
- (3) C10.000

Quem adota a estratégia (a) deverá rejeitar a premissa (1) ou a premissa (2). Deixemos (2) de lado por enquanto. Rejeitar a premissa (1) implica em rejeitar que uma pessoa com 0 fios de cabelo na cabeça seja careca. De partida, isso nos parece inaceitável. A estratégia (b) envolve rejeitar a Indução Matemática. Este caminho não parece mais promissor do que o anterior; sem contar que na versão condicional o paradoxo é derivado com Modus Ponens e não com Indução Matemática. Os argumentos sorites envolvem regras de inferência básicas demais para serem rejeitadas. Por fim, aceitar a conclusão nem de longe é uma opção animadora.

A opção que nos resta é retornar à estratégia (a) e rejeitar a premissa (2), isto é, rejeitar o princípio de tolerância para “careca”. Por um lado, o princípio não é tão intuitivo quanto

a afirmação de que uma pessoa com 0 fios de cabelo na cabeça é careca. Por outro, rejeitar o princípio de tolerância não é tão drástico quanto rejeitar o Modus Ponens, Indução Matemática ou aceitar a conclusão do sorites. Mesmo que haja uma ou outra intuição favorecendo o princípio, sua rejeição certamente requer um sacrifício menor do que qualquer outra opção.

A rejeição do princípio de tolerância também nos fornece uma solução para as outras versões do paradoxo. Na versão condicional, a solução consistiria novamente em rejeitar uma das premissas. Lembre que cada premissa condicional é uma instância do princípio de tolerância. Uma vez que rejeitamos o princípio, podemos rejeitar alguma de suas instâncias e, conseqüentemente, resolver a versão condicional do sorites. Na versão da linha desenhada, a solução será aceitar a conclusão. Nessa versão, a conclusão do argumento é precisamente a negação do princípio de tolerância. Dado que rejeitamos o princípio, podemos aceitar a sua negação e, conseqüentemente, resolvermos o paradoxo.

Em suma, a rejeição do princípio de tolerância é a opção inicialmente mais plausível e nos permite resolver todas as versões acima do paradoxo. Entretanto, essa opção enfrenta um sério problema.

2.6. O problema da precisificação

Pense em predicados como “careca”, “monte”, “alto”, “grande”, etc. Esses predicados não parecem estabelecer divisões precisas entre os objetos aos quais se aplicam e os objetos aos quais não se aplicam. Na verdade, é uma propriedade essencial dos predicados vagos que eles não estabeleçam qualquer divisão deste tipo; é isto que os diferencia dos predicados precisos como “ser par” ou “ser ímpar”. De fato, predicados vagos são essencialmente imprecisos.

Pode ajudar pensar em termos de sequências. Imagine uma sequência como segue:

$$\langle a_0, a_1, a_2, \dots, a_{5.000}, a_{5.001}, \dots, a_{10.000} \rangle$$

Primeiro, suponha que cada “ a ” na sequência seja o nome de uma pessoa, e o número na direita inferior represente o número de cabelos que a pessoa tem na cabeça. O predicado “careca” se aplica a a_0 , e não se aplica a $a_{10.000}$. Mas não há uma divisão precisa entre os casos nos quais “careca” se aplica e os casos nos quais não se aplica. Se houvesse tal divisão,

“careca” seria um predicado preciso e não um predicado vago. O mesmo vale para qualquer outro predicado vago. Imagine agora que “ $a_{10.000}$ ” nomeia uma aglomeração de 10.000 grãos de trigo que claramente formam um monte, “ $a_{9.999}$ ” nomeará o resultado de extrairmos um único grão de $a_{10.000}$ e assim sucessivamente. O predicado “Monte de trigo” se aplica a $a_{10.000}$ e não se aplica a a_1 , mas não há uma divisão precisa entre os casos nos quais o predicado se aplica e os casos nos quais não se aplica. Se houvesse essa divisão, “monte de trigo” seria preciso e não vago.

Galeno certa vez disse que não conhecia nada “pior e mais absurdo do que [a tese de que] o ser e o não-ser de um monte é determinado por um [único] grão” (*On Medical Experience* XVII, p.116, tradução minha). Apesar do provável exagero e de expressar-se com uma tese metafísica, podemos assumir que Galeno também aceitaria a dificuldade da ideia de que o predicado “monte” estabelece uma divisão precisa entre as coisas às quais se aplica e as coisas às quais não se aplica. Poucos recusarão o aspecto altamente contraintuitivo da tese de que predicados vagos cortam o mundo de forma *cirúrgica*. Mas o problema aqui não é meramente que essa tese é contraintuitiva. O principal problema é que isso implica que predicados vagos são, no fim das contas, precisos e não vagos.

Um predicado é preciso se existe uma fronteira entre as coisas às quais ele se aplica e as coisas às quais ele não se aplica. Uma propriedade essencial dos predicados vagos é que eles são imprecisos. Um predicado é impreciso se não admite tal fronteira, nenhuma linha fazendo um corte entre os casos nos quais o predicado se aplica e os casos nos quais não se aplica. Como disse Wright, se um predicado é impreciso, então “nenhuma tal distinção pode ser traçada entre os casos aos quais é definitivamente correto aplicar o predicado e os casos de qualquer outro tipo.” (1975, p.330, tradução minha).

Terence Horgan (1994, p.162) chamou de *o problema da precisificação* ao problema da explicar a vagueza sem implicar que os predicados vagos sejam predicados precisos; isto é, sem precisificar os predicados vagos. Horgan argumentou plausivelmente a favor de que esse problema é mais difícil do que inicialmente parece. Eis a principal dificuldade. Vimos que a estratégia inicialmente mais plausível para resolver o paradoxo sorites é rejeitar o princípio de tolerância. Qualquer uma das outras estratégias é muito mais difícil de engolir. A rejeição do princípio de tolerância, contudo, implica na precisificação dos predicados vagos. Em outras palavras, ao rejeitar o princípio implicamos que os predicados vagos são predicados precisos.

Considere o princípio de tolerância aplicado ao predicado “careca”: $\forall n (C_n \rightarrow C_{n+1})$. A rejeição desse princípio nos compromete com a tese de que ao menos uma das instâncias da fórmula universalmente quantificada não é verdadeira. Suponha que esta instância seja a

condicional $C5.000 \rightarrow C5.001$. Nesse caso, a antecedente dessa condicional é verdadeira, mas a consequente não. Por outras palavras, uma pessoa com 5.000 fios de cabelo na cabeça é careca e uma pessoa com 5.001 fios de cabelo na cabeça não é careca. Resultado: existe uma fronteira precisa entre as pessoas que são carecas e as que não são. Aquelas que possuem no máximo 5.000 fios de cabelo na cabeça são carecas, os outros não. Ora, isto significa que o predicado “careca” é um predicado preciso, um predicado que traça uma linha entre aqueles objetos aos quais ele se aplica e aqueles objetos aos quais não se aplica. O problema é que se “careca” é preciso, então não é vago. Em outras palavras, rejeitar a tolerância de “careca” nos compromete com rejeitar a sua vagueza. O mesmo ocorreria para qualquer outro predicado vago.

Este problema foi parcialmente antecipado quando vimos a razão pela qual seria implausível rejeitar alguma das premissas condicionais da versão condicional do paradoxo. Ao considerarmos o caso de “careca”, vimos que a rejeição de alguma condicional nos comprometia com uma divisão precisa entre os que são carecas e os que não são carecas. Esta consequência era muito contra-intuitiva. Agora podemos perceber que o problema vai além do fato de a consequência ser contra-intuitiva. O principal problema é que rejeitar alguma condicional nos compromete em rejeitar a própria vagueza dos predicados vagos. Seja lá o que for que os predicados vagos sejam, uma coisa parece clara: eles não são precisos, não estabelecem um corte preciso entre os objetos aos quais eles se aplicam e os objetos aos quais eles não se aplicam.

Por fim, é por isso que não podemos simplesmente resolver a versão da linha desenhada do paradoxo sorites aceitando a sua conclusão ($\exists n (C_n \wedge \neg C_{n+1})$), que é logicamente equivalente à negação do princípio de tolerância ($\neg \forall n (C_n \rightarrow C_{n+1})$). O que $\exists n (C_n \wedge \neg C_{n+1})$ afirma é que há um número n tal que uma pessoa com n fios de cabelo na cabeça é careca e uma pessoa com $n+1$ fios na cabeça não é careca. Por outras palavras, há uma fronteira precisa entre os carecas e os que não são carecas. Novamente, o resultado é que o predicado “careca” é preciso e não vago.

Existem muitas teorias que tentam resolver o sorites rejeitando o princípio de tolerância. Cada uma delas enfrenta o problema da precisificação, e não é claro que consigam escapar dele. Tomando o caso de “careca” como exemplo, o problema é que as teorias aceitam que em algum sentido existe um número n tal que C_n é verdadeiro e C_{n+1} não é verdadeiro. Teorias dos graus de verdade aceitarão a existência de um número n tal que C_n é verdadeiro em grau 1 e C_{n+1} não é verdadeira em grau 1 (é verdadeira em grau menor do que 1). Supervaloracionistas aceitarão que existe um número n tal que C_n é verdadeiro para toda precisificação e C_{n+1} não é verdadeiro para toda precisificação. Contextualistas aceitarão a existência de uma fronteira entre carecas e não carecas e tentarão explicar a vagueza –

ou pelo menos a aparente tolerância dos predicados vagos – apelando a um *requisito de similaridade* que, grosso modo, diz que a fronteira dos predicados vagos nunca está onde nós *a procuramos*. Em suma, essas e outras teorias que rejeitam o princípio de tolerância têm dificuldades em escapar do problema da precisificação.

O problema da precisificação levou alguns filósofos a procurarem saídas alternativas para o sorites. Uma saída é aceitar o princípio de tolerância e a consequente conclusão de que predicados vagos são em algum sentido incoerentes. Esta saída é adotada pelo próprio Horgan, e também por Wright (1975, 1976) e Eklund (2005). Trata-se de uma resposta de partida indesejável, dado que é difícil de acreditar que nossos predicados vagos – que provavelmente são a maior parte de nossos predicados – sejam incoerentes. Entretanto, apesar de serem inicialmente indesejáveis, talvez as soluções desse tipo sejam as únicas que evitem o problema da precisificação, explicando a vagueza ao invés de eliminá-la. Em todo caso, temos um difícil problema pela frente. Por um lado, a rejeição do princípio de tolerância parece a saída mais plausível para o paradoxo sorites. Por outro, a rejeição do princípio parece implicar na precisificação dos predicados vagos.

2.7. Critérios de adequação

Teorias da vagueza devem explicar o fenômeno da vagueza. É muitíssimo desejável que o façam de modo a resolver o paradoxo sorites. Afinal, é este paradoxo que mais chama a nossa atenção para o fenômeno. É também desejável que as teorias não impliquem na incoerência dos predicados vagos. A conclusão de que os predicados “careca”, “alto”, “grande” e tantos outros são incoerentes nos soa não apenas como surpreendente, mas também absurda. Aliás, é em parte por isso que o paradoxo é um paradoxo. Pudéssemos aceitar facilmente a sua conclusão, não se trataria de um paradoxo. Por fim, é desejável que as teorias não impliquem que predicados vagos são predicados precisos. Não fosse assim, o problema da precisificação sequer seria um problema.

Isto tudo nos leva a três critérios de adequação para uma teoria ideal da vagueza; três critérios que esperamos que uma teoria que explique a natureza da vagueza satisfaça:

- (a) *O critério do sorites*: resolver o paradoxo sorites.
- (b) *O critério da coerência*: não implicar que predicados vagos são incoerentes.

(c) O critério da precisificação: não implicar que predicados vagos são precisos.

Repare que estes critérios não surgem de teorização sofisticada sobre vagueza, mas aparecem tão logo começamos a pensar sobre o sorites e os modos iniciais de resolvê-lo. Além disso, cada um deles esteve por vezes explicitamente e por vezes implicitamente presente ao longo da discussão sobre a vagueza, desde a antiguidade grega e romana. O caso dos dois primeiros critérios é mais intuitivo, de modo que suponho que seja difícil discordar que (a) e (b) sejam realmente critérios de adequação que uma teoria ideal da vagueza deveria satisfazer. O último pode ser visto com maior desconfiança.

De fato vagueza parece ser um fenômeno sobre fronteiras e, mais especificamente, sobre uma aparente inexistência de fronteiras. Além disso, se a postulação de uma fronteira não nos parecesse absurda, teríamos um modo imediato de resolver o paradoxo: postular a existência de fronteiras. A intuição de que os predicados vagos não cortam o mundo de forma precisa já tem uma longa história. Como vimos logo acima, Galeno descrevia a ideia de um corte preciso entre montes e não-montes como a coisa mais absurda que conhecia. Entre os estóicos já era considerado intuitivo que não exista um n tal que uma pessoa com n fios de cabelo na cabeça é careca e uma pessoa com $n+1$ fios não. Não é uma surpresa que a rejeição dessa intuição seja vista como um problema para qualquer teoria, como ressaltou Horgan ao apresentar o problema da precisificação. Alguns encaram esse problema muito seriamente, acreditando que violar o critério de precisificação é uma condição suficiente para a rejeição de uma teoria. Outros pensam que, no fim das contas, a violação do critério de precisificação pode ser tolerada. Mesmo entre os últimos, contudo, é normalmente aceito que o ônus recai sobre quem o viola, isto é, a violação do critério de precisificação requer justificação.² Em todo caso, na próxima seção apresentarei uma razão extra pela qual devemos levar o critério de precisificação muito seriamente.

Não é minha sugestão aqui que a violação de qualquer um dos critérios acima é inaceitável. Ao dizer que eles são critérios para uma *teoria ideal da vagueza* quero apenas dizer que é desejável que uma teoria não os viole. A violação de qualquer um desses critérios será encarada como uma desvantagem para a teoria, assim como a não violação será encarada como uma vantagem. Em todo caso, é no mínimo difícil satisfazer conjuntamente aos três critérios, e do que foi dito está aberta a possibilidade de que sequer possamos satisfazê-los conjuntamente.

2 - Além de Horgan (1994), alguns exemplos do primeiro grupo são Wright (1975,1976), Tye (1994, p.193), Fodor & Lepore (1996, p.523), Keefe (2000), Eklund (2005, p.41) etc. Alguns exemplos do segundo grupo são Fara (2000), Priest (2003, p. 11), etc. Não estou dizendo que esses filósofos afirmem explicitamente aceitar o critério da precisificação. Muitos deles não usam essa expressão, nem afirmam isso diretamente. Mas os do primeiro grupo pensam ser inaceitável estabelecer uma fronteira entre os casos nos quais um predicado vago se aplica e os casos nos quais não. Os do segundo pensam que o estabelecimento de uma fronteira do tipo consiste pelo menos numa desvantagem que precisa ser motivada para ser aceita.

2.8. O problema fundacional da precisificação

Nesta seção apresento uma razão a mais pela qual devemos tomar o critério de precisificação seriamente. Já vimos uma razão para isso. A imprecisão parece ser uma propriedade essencial dos predicados vagos; se um predicado F é vago, então não pode ser o caso que F é preciso. Qualquer teoria que afirme que predicados vagos são precisos está, no fim das contas, afirmando que não são vagos de todo em todo. No entanto, acredito que há ainda outra razão para levarmos o critério de precisificação a sério. De forma direta: teorias que violam o critério de precisificação terão problemas em explicar como a fronteira precisa dos predicados vagos é determinada. Chamarei a este problema de *o problema fundacional da precisificação*.

Muitos dos predicados de nossa linguagem natural se aplicam a algumas coisas. Geralmente, há coisas às quais eles se aplicam e coisas às quais não se aplicam. Considere o predicado “careca”. Esse predicado se aplica a algumas pessoas e não se aplica a outras. Costumamos pensar que “careca” se aplica a pessoas com 0 fios de cabelo e não se aplica a pessoas com 10.000 fios de cabelo na cabeça. É claro que a aplicação do predicado não é feita por mágica. Não proferimos predicados e, magicamente, eles passam a aplicar-se ou não às coisas. É preciso ter algum mecanismo determinando o sucesso ou insucesso da aplicação. O papel de uma teoria da referência dos predicados é explicar qual mecanismo é esse e como ele funciona, garantindo o sucesso ou insucesso da aplicação dos predicados.

Se predicados vagos são precisos, então há uma fronteira entre os itens aos quais eles se aplicam e os itens aos quais eles não se aplicam. Isso implica que o mecanismo referencial desses predicados deve determinar essa fronteira precisa. É legítimo perguntar como isso ocorre. A referência não pode ser determinada por mágica, tem de haver um mecanismo garantindo que os predicados tenham a fronteira que têm. Quem acredita que esses predicados são precisos, terá o problema de explicar como a sua suposta fronteira precisa é determinada. Essa é uma tarefa bastante complicada.

O que torna especialmente difícil explicar como predicados vagos adquirem fronteiras precisas é que aparentemente há várias fronteiras com igual direito ao cargo de fronteira precisa de um predicado vago. Imagine que alguém sustente que o predicado “careca” se aplica a pessoas com no máximo n fios de cabelo na cabeça, e não se aplica a ninguém mais. Você poderia legitimamente perguntar como é que a fronteira foi determinada em n ao invés de, por exemplo, $n+1$ ou $n-1$. Como é que o mecanismo referencial de “careca”

garantiu que este predicado se aplique a pessoas com no máximo n – ao invés de $n+1$ ou $n-1$ – fios de cabelo na cabeça? É difícil perceber como o mecanismo referencial das palavras vagas seleciona uma fronteira específica dentre todas as candidatas.

Em suma, quem defende que há uma fronteira entre “careca”/“não-careca”, “monte”/“não-monte”, “alto”/“não-alto”, “magro”/“não-magro” (etc.) terá de lidar com o problema de como tal fronteira é determinada. Suponho que isso não seja muito controverso.

Por fim, é interessante ilustrar brevemente como o problema pode surgir. Não há dificuldades em pensar em como poderíamos determinar uma fronteira precisa para alguns predicados vagos. Isso poderia ser feito, por exemplo, por meio de estipulação. Poderíamos estipular que o predicado “careca” se aplica a pessoas com no máximo n fios de cabelo na cabeça e não se aplica a qualquer outro, poderíamos estipular que se aplica a pessoas com no máximo o mesmo número de cabelos que João, etc. É difícil acreditar, contudo, que a fronteira de “careca” seja de fato determinada por estipulação. Apesar de tais estipulações estarem por vezes disponíveis, nosso uso de “careca” e outros predicados vagos raramente é acompanhada por estipulações de fronteiras. Nosso problema é saber como a fronteira de “careca” (e de outros predicados vagos) é de fato determinada.

Eis um modo de responder ao problema. Comece notando que o predicado “cavalo” se aplica às coisas que são cavalos, e somente a elas. Um modo natural de expressar isso é dizendo que o predicado “cavalo” se aplica às coisas que possuem a propriedade de *ser cavalo*, e a nada mais. Todos os cavalos, e somente os cavalos, possuem a propriedade de *ser cavalo*, e é essa propriedade que determina a extensão desse predicado. Do mesmo modo, os carecas, e somente os carecas, possuem a propriedade de *ser careca* e é essa propriedade que determina a extensão de “careca”. A ideia central aqui é que o mecanismo referencial de “careca” determina que o predicado se aplica a todos, e apenas a todos, os que possuem a propriedade de *ser careca*.

O principal problema dessa perspectiva é que há muitas propriedades candidatas ao cargo de propriedade de *ser careca* e a escolha entre elas é arbitrária. Por exemplo, há a propriedade de ter no máximo 0 fios de cabelo na cabeça, a propriedade de ter no máximo 1 fio de cabelo na cabeça, a propriedade de ter no máximo 2 fios de cabelo na cabeça, e assim por diante. Por que uma delas seria uma candidata melhor do que as outras ao cargo de propriedade de *ser careca*? Há várias propriedades aparentemente com igual direito ao cargo de *propriedade que determina a fronteira de “careca”*. A plausibilidade da tese acima depende de justificarmos que alguma das propriedades candidatas seja especial em relação às outras.

Uma saída pode ser procurada em uma teoria causal da referência dos predicados.

Grosso modo, teorias causais dizem que (um uso de) um predicado F se aplica a um objeto x se existe um elo causal apropriado entre (objetos com a mesma natureza de) x e (o uso de) F . As coisas supostamente funcionam bem para predicados como “água”. Em algum momento alguém apontou para um exemplar de água e introduziu o predicado “água” para falar sobre aquele exemplar e tudo o que tiver a *mesma natureza* que ele. Desde então, o predicado foi difundido de pessoa para pessoa, formando uma longa cadeia causal, com origem naquele exemplar batizado, mas que nos leva a tudo que tenha a mesma natureza que ele. A natureza daquele primeiro exemplar de água era a sua composição química: H_2O . O predicado “água”, usado por um membro dessa cadeia, aplica-se àquele primeiro elemento e a tudo que tenha a mesma composição química H_2O .

Repare que em teorias causais a introdução de um predicado envolve dois elementos: o *elemento ostensivo* e o *elemento de natureza* (Devitt & Sterelny, 1999:88). Para introduzir o predicado “água”, temos de estar em contato com um exemplar de água (esse é o elemento ostensivo). Mas isso não basta. Não introduzimos “água” para falar apenas daquele exemplar específico, mas também de todo exemplar que *tenha a mesma natureza* que o primeiro. Assim, uma condição necessária para as coisas correrem bem é que exista uma natureza comum a todas as coisas que chamamos “água”.

Algo similar ocorreria com “careca”. Em algum momento alguém introduziu “careca” para falar de um indivíduo específico e todos aqueles que partilham certa natureza com ele. Essa natureza em comum é a propriedade de *ser careca*. O predicado então foi difundido de falante para falante até nos alcançar. Quando usado por um membro dessa cadeia, o predicado se aplica a todos, e apenas a todos, aqueles que possuem a propriedade de *ser careca*. Essa tese será implausível para aqueles que rejeitam tese metafísica de que há uma propriedade de *ser careca*. Mas o ponto para o qual quero chamar a atenção é que o problema anterior permanece intacto.

Há muitos candidatos ao cargo de propriedade de *ser careca*, e não há um específico que mereça mais o cargo do que todos os outros. Para que a teoria causal possa ser usada por aqueles que violam o critério de precisificação, tem de haver uma propriedade específica que cumpre um papel especial na determinação da fronteira do predicado. Mas é difícil justificar que exista alguma propriedade que cumpra esse papel. Considere o seguinte:

A propriedade de ter no máximo $n-1$ fios de cabelo na cabeça

A propriedade de ter no máximo n fios de cabelo na cabeça

A propriedade de ter no máximo $n+1$ fios de cabelo na cabeça

Não importa qual o valor de n , nós poderemos sempre perguntar o que torna uma das propriedades acima mais relevante do que as outras duas para a determinação da fronteira de “careca”. Analogamente, aqueles que apelam à teoria causal teriam que explicar o que faz com que uma delas ocupe o papel especial na cadeia causal de usos de “careca”. Precisamos de uma razão para acreditar que exista alguma delas que seja especial em relação às outras. O que aparenta ser fácil para o predicado “água” pode ser bastante complicado para “careca” e outros predicados vagos.

Quem viola o critério de precisificação tem de aceitar que predicados vagos são precisos. Quem aceita que predicados vagos são precisos tem de explicar como a sua fronteira precisa é determinada. Portanto, quem viola o critério de precisificação tem de explicar como a fronteira precisa dos predicados vagos é determinada. Como illustrei acima, esta última tarefa pode ser mais difícil do que parece. Não se trata de defender que é um problema insolúvel, mas de reconhecer sua existência e dificuldade.

Em resumo, os três critérios de adequação acima surgem tão logo começamos a pensar sobre o sorites e as estratégias mais gerais de solução. Todos os três critérios são largamente intuitivos, e nenhum deles pode ser facilmente rejeitado. Seria desejável que as teorias da vagueza os satisfizessem, supondo que isso seja possível. Isto é suficiente para mostrar a plausibilidade de tomá-los como critérios de adequação para uma teoria ideal da vagueza.

3. Intuições sobre vagueza

Nossa atenção ao fenômeno da vagueza foi majoritariamente voltada para o paradoxo sorites. Todos os três critérios de adequação apontados para uma teoria da vagueza surgiram no contexto da apresentação do sorites. Vimos que uma teoria ideal da vagueza deve explicar a natureza da vagueza de forma a fornecer uma resposta ao sorites (critérios sorites), e que isso deveria ser feito sem implicar que os predicados vagos sejam em algum sentido incoerentes (critério da coerência) ou precisos (critério da precisificação). Certamente que o fenômeno da vagueza e o paradoxo sorites estão intimamente ligados, e resolver o paradoxo é uma forte motivação para procurarmos por uma teoria da vagueza.

Apesar disso, a discussão sobre a vagueza vai além da discussão sobre o paradoxo sorites. Como em grande parte das discussões filosóficas, as discussões sobre vagueza começam com caracterizações iniciais do fenômeno a ser explicado. O problema é que o fenômeno

da vagueza tem sido inicialmente caracterizado de muitas formas diferentes, e sequer é claro que sejam caracterizações consistentes entre si. Isso dificulta a discussão, pois gera o risco de que diferentes teorias sejam direcionadas para diferentes fenômenos. Além disso, mesmo que todas as caracterizações iniciais sejam direcionadas para um mesmo fenômeno, não é nada claro qual delas está correta. Talvez algumas caracterizações iniciais capturem apenas as *bordas* do fenômeno da vagueza, deixando seu *centro* de fora. De todo modo, podemos tomar cada uma dessas caracterizações como representando intuições comumente aceitas acerca do que está envolvido no fenômeno da vagueza.

Digamos que uma intuição é tratada como *fundamental* com respeito à vagueza quando se considera que qualquer teoria da vagueza deve se acomodar a ela. As intuições *mais fundamentais* são aquelas em termos das quais as outras são explicadas. Qualquer teoria da vagueza terá de dizer algo sobre a organização das nossas intuições mais comuns sobre o tema. Alguma delas pode ser tomada como (a mais) fundamental? Quais? Quais serão aceitas apenas como representando aspectos menos importantes do fenômeno? Alguma delas será rejeitada como uma intuição equivocada? Etc.

Como era de se esperar, não é fácil escolher qual ou quais delas captam mais adequadamente o fenômeno da vagueza. Minha própria escolha é guiada pelos critérios de adequação acima. Acredito que a *intuição da arbitrariedade da fronteira* é a mais fundamental, principalmente porque nos fornece o caminho para uma teoria que satisfaz os três critérios de adequação para uma teoria ideal; mas isso é assunto para outro artigo. Nesta seção, apresento uma lista de intuições comuns sobre a natureza da vagueza.

Antes, algumas observações importantes. Nenhuma intuição listada receberá uma formulação rigorosa, e isto é proposital. As intuições ganham formulações rigorosas apenas já no seio das teorias da vagueza; e uma mesma intuição pode receber tratamentos diferentes em diferentes teorias. A exposição aqui tem de ser suficientemente genérica para permitir que a mesma intuição possa ser interpretada diferentemente por diferentes teorias. Por fim, formulo as intuições como intuições sobre a linguagem, isto é, sobre *predicados vagos*. Mas nada impede que sejam formuladas como intuições sobre o mundo, isto é, sobre *propriedades vagas*. Passemos então à lista.

3.1. Uma lista de intuições

(i) *A Intuição da Ausência de Fronteira*

Já vimos essa intuição ao longo de nossa discussão. Trata-se da tese de que predicados

vagos não têm fronteira precisa. Não há uma fronteira precisa entre os casos nos quais o predicado se aplica e os casos nos quais o predicado não se aplica. Essa intuição esteve presente – se ou não foi afirmada explicitamente – desde o princípio da discussão sobre o sorites. Horgan (1994) e Wright (2009, p.530) tomaram esta intuição como a mais fundamental sobre vagueza; Keefe (2000, p.153), Tye (1994, p.193) no mínimo a consideraram fundamental.

(ii) *Intuição da Questão de Fato*

Não existe uma questão de fato acerca de onde é a fronteira de um predicado vago. Não é nada claro o que isso signifique. Pode ser entendido como um modo alternativo de afirmar que predicados vagos não têm fronteira. Mas também pode ser interpretado de outros modos. Uma possível interpretação é que isso expressa indeterminação da fronteira. Por exemplo, é argumentável que não exista uma questão de fato acerca de quem será o próximo presidente da Argentina. Dizer isto é dizer que é agora *indeterminado* quem será o próximo presidente da Argentina. De modo análogo, podemos entender “não existe uma questão de fato sobre onde é a fronteira dos predicados vagos” como significando que a fronteira dos predicados vagos é em algum sentido indeterminada. (Importante, o sentido de *indeterminado* relevante para a vagueza provavelmente será diferente daquele relevante para discussões sobre o futuro ser indeterminado). Field (2003, p. 457) claramente toma esta intuição como a mais fundamental; Keefe (2000, p.153) e Tye (1994, p.193) a tomam como fundamental.

(iii) *A Intuição da Tolerância*

Também já considerada, trata-se da tese de que predicados vagos são tolerantes a mudanças suficientemente pequenas. Por outras palavras, as diferenças muitíssimo pequenas não fazem diferença para a aplicação do predicado. O acréscimo de um único fio de cabelo não torna um careca em não-careca, assim como a retirada não torna um não-careca em careca; o acréscimo de um único grão de trigo não torna um não-monte de trigo em um monte, assim como a retirada não torna um monte em um não-monte. Essa intuição também esteve presente – se ou não afirmada explicitamente – ao longo de toda a discussão sobre vagueza. Foi C. Wright (1975, 1976), contudo, quem desenvolveu a noção

de *tolerância*, tomando-a explicitamente como fundamental naquele momento.

(iv) *A Intuição dos Casos Fronteira*

Esta intuição ainda não foi considerada, mas tem uma enorme importância na discussão sobre vagueza. Basicamente, a ideia é que predicados vagos admitem casos fronteiras. Em uma definição neutra, um caso fronteira de um predicado F é um caso em relação ao qual a aplicação de F não é clara. Uma pessoa com 0 fios de cabelo claramente é careca, enquanto uma pessoa com 10.000 fios de cabelo claramente não é careca. Mas o que dizer de uma pessoa com 80% do couro cabeludo coberto de cabelo? Nesse caso, não é claro se o predicado “careca” se aplica ou não; trata-se de um caso fronteira de “careca”. Muitos filósofos acreditam que a admissão de casos fronteira é uma característica fundamental, se não a mais fundamental, dos predicados vagos. O exemplo clássico é Kit Fine (1975), outro exemplo M. Richard (2009, p. 465-467).

(v) *A Arbitrariedade da Fronteira*

Há uma variedade de fronteiras que podemos atribuir aos predicados vagos. Já vimos isso na seção anterior. Podemos estabelecer diferentes fronteiras para “careca”, mas qualquer fronteira que determinemos será arbitrária. A pergunta “por que esta fronteira e não alguma das outras?” surge tão logo tentemos escolher por uma. Mas nenhuma resposta vem à mente. Geralmente, isso é expressado dizendo-se que *toda precisificação de um predicado vago é igualmente arbitrária*. Horgan (1994) e Soames (1998, p. 206) consideram essa intuição fundamental (mas talvez não a mais fundamental). Acredito que essa é a intuição mais fundamental sobre vagueza.

(vi) *A Incognoscibilidade da Fronteira*

É impossível para nós humanos conhecermos onde está a fronteira de um predicado vago. Não é possível para nós sabermos onde é a fronteira exata entre carecas/não-carecas, montes/não-montes, altos/não-alto, etc. Talvez a explicação para isso seja que não exista tal fronteira. Mas do mero fato de que não podemos conhecer não se segue que não exista. Se

ou não existe uma fronteira, o fato é que intuímos fortemente que não podemos conhecê-la. Timothy Williamson (1992, 1994, 1997, etc.) tem gastado muita tinta para defender que essa é a intuição fundamental sobre vagueza.

Moral da História

De um modo direto, nossa lista é a seguinte:

- (i) *A Intuição da Ausência de Fronteira*
- (ii) *Intuição da Questão de Fato*
- (iii) *A Intuição da Tolerância*
- (iv) *A Intuição dos Casos Fronteira*
- (v) *A Arbitrariedade da Fronteira*
- (vi) *A Incognoscibilidade da Fronteira*

Estas não são as únicas caracterizações iniciais sobre vagueza, e não representam as únicas intuições sobre o que está envolvido no fenômeno.³ Mas acredito que a lista acima seja justa o bastante, incluindo as intuições mais comumente aceitas sobre vagueza. (Compare, por exemplo, com Ronzitti (2011, p.v), que inclui apenas três intuições na lista).

Meu principal objetivo aqui é chamar a atenção para dois pontos. Primeiro, não temos qualquer garantia de que as intuições (i)-(vi) podem ser interpretadas umas em termos das outras. Pior, até onde sabemos está aberta a possibilidade de que algumas delas sequer sejam consistentes entre si. Segundo, existe o problema de organizar estas intuições. Por *organizar* quero dizer (i) determinar quais das intuições acima aceitamos e quais rejeitamos e (ii) determinar, dentre aquelas que aceitamos, quais são mais fundamentais e quais são menos fundamentais.

3 - Wright (1994, p.138) certa vez defendeu que uma intuição fundamental sobre predicados vagos é que eles aceitam alguns casos de desacordo permitido. Seja *F* um predicado vago, e *Fa* um caso em relação ao qual o desacordo é permitido. Então duas pessoas podem discordar acerca de se *a* é ou não *F* sem que nenhum dos dois esteja errado. Apesar de concordar com Wright que predicados vagos realmente têm essa característica, não me é claro que isso possa ser tomado como algo intuitivo sobre predicados vagos. (Aliás, você já deve ter percebido que Wright não parece muito seguro sobre o que considera ou não como intuição fundamental sobre vagueza).

3.2. Como não refutar uma Teoria da Vagueza

Uma importante lição da seção anterior é que devemos ter cautela com um tipo específico de objeções a teorias da vagueza. Seja T uma teoria que explica a natureza dos predicados vagos. Alguém pode argumentar que T é incorreta ou por ser muito restrita ou por ser muito ampla. No primeiro caso, forcemos um exemplo de predicado vago que a teoria implica ser não-vago. No segundo, fornecemos um exemplo de predicado não-vago que a teoria implica ser vago.

Argumento por contra-exemplo positivo: T é restrita demais: existe uma (ou um tipo de) expressão E que é tal que: E é vago e T implica que E não é vago.

Argumento por contra-exemplo negativo: T é ampla demais: existe uma (ou um tipo de) expressão E que é tal que: E não é vago e T implica que E seja vaga.

Esse tipo de objeção é legítimo, mas no caso da vagueza requer um cuidado especial. Como vimos, o fenômeno da vagueza é associado a muitas intuições diferentes, e sequer é claro que sejam intuições consistentes entre si. No fim das contas – e eu de fato penso que este é o caso – pode ser que toda teoria da vagueza tenha de rejeitar uma ou mais das intuições acima. Se isso é assim, então nenhuma teoria se acomodará perfeitamente a todas as intuições mais comuns sobre a natureza da vagueza. Ao fornecermos um argumento por contra-exemplo, devemos tomar o cuidado de não apelar a alguma intuição que já foi adequadamente rejeitada.

Eis como a coisa toda pode ocorrer. Suponha que uma teoria T rejeite a intuição I. Agora, podemos criar um contra-exemplo positivo da seguinte forma. Primeiro, crie um predicado artificial que satisfaça a intuição I. Em seguida, reivindique que ele é vago, mas que T implica que é não-vago. Feito. Podemos criar um contra-exemplo negativo do seguinte modo. Primeiro, crie um predicado artificial que satisfaça as condições estipuladas pela teoria T, mas não se adéque a I. Em seguida, reivindique que ele é não-vago e T implica que seja vago. Feito.

O problema é que a referência à intuição pode estar apenas implícita. Alguém pode dizer simplesmente que um predicado vago não é reconhecido como vago por T, sem explicitar em qual sentido o predicado é tomado como vago ou qual intuição sobre a vagueza está em jogo. Do mesmo modo, alguém pode argumentar que um predicado não-vago é tratado como vago por T, sem tornar explícito qual intuição está em jogo. Dado o que vimos,

qualquer argumento desse tipo deve ser visto com suspeita.

4. O problema da vagueza

Chegamos finalmente à formulação final do problema da vagueza:

Problema da vagueza: o problema de explicar os predicados/expressões vagas de forma a resolver o paradoxo sorites, e sistematizar as intuições (i)-(vi); se possível, satisfazendo conjuntamente todos os três critérios de adequação (sorites, coerência e precisificação).

A formulação acima do problema da vagueza nos fornece um ponto de partida para começarmos a pensar em soluções. De fato, com ela já temos algumas diretivas sobre o que procurar em uma solução da vagueza (a satisfação conjunta dos três critérios e a sistematização de um conjunto de intuições) e de quais dificuldades podemos encontrar pelo caminho (a dificuldade de satisfazer conjuntamente os três critérios, a ampla variedade de intuições para serem explicadas). Certamente que se trata apenas de um começo. No entanto, o primeiro passo para resolver um problema filosófico é torná-lo mais claro. Espero ter contribuído nesse sentido.

Referências

Burnnycat, M. F. (1982). Gods and heaps. In: M. Schofield, & M. C. Nussbaum, *Language and Logos* (pp. 315-338). Cambridge University Press.

Devitt, M., & Sterelny, K. (1999). *Language and Reality: an introduction to the philosophy of language*. Blackwell.

Dummett, M. (1975). Wang's Paradox. *Synthese* 30 (30), 301–324.

Eklund. (2005). What Vagueness Consists in. *Philosophical Studies* (125), 27–60.

Empirico, S. *Outlines of Pyrrhonism*. (L. R. G. Bury, Trad.) Harvard University Press.

- Fara, D. G. (2000). Shifting Sands: An Interest-Relative Theory of Vagueness. *Philosophical Topics*, 28 (1).
- Field, H. (2003). No fact of the matter. *Australasian Journal of Philosophy* (81), 457–480.
- Fine, K. (1975). Vagueness, Truth and Logic. *Synthese* (30), 265-300.
- Fodor, J. A., & Lepore, E. (1996). What Cannot be Evaluated Cannot be Evaluated and it Cannot be Supervalued Either. *The Journal of Philosophy*, 96 (10), 516-535.
- Galeno, C. *On Medical Experience* (Vol. Three Treatises on The Nature of Science). (R. Walzer, & M. Frede, Trans.) Hackett Publishing Company.
- Horgan, T. (1994). Robust Vagueness and the Forced-March Sorites Paradox. *Philosophical Perspectives*, 8 (Logic and Language), 159-188.
- Hyde, D. (2011). The sorites paradox. In: G. Ronzitti, *Vagueness: A guide* (pp. 1-17). Springer Science & Business Media.
- Keefe, R. (2000). *Theories of Vagueness*. Cambridge University Press.
- Laërtios, D. *Vidas e Doutrinas dos Filósofos Ilustres*. (M. d. Cury, Trad.) Editora UnB.
- Mignucci, M. (1993). The Stoic Analysis of the Sorites. *Proceedings of the Aristotelian Society* (93), 231–245.
- Priest, G. (2003). A site for sorites. In: J. Beall, *Liars and Heaps* (pp. 09-23). Oxford University Press.
- Richard, M. (2009). Indeterminacy and Truth Value Gaps. In: R. Dietz, & S. Moruzzi, *Cuts and Clouds*. Oxford University Press.
- Ronzitti, G. (2011). Vagueness and... In: G. Ronzitti, *Vagueness: A guide* (pp. v-x). Springer Science & Business Media.
- Sainsbury, R. M., & Williamson, T. (1998). Sorites. In: B. Hale, & C. Wright, *A Companion to Philosophy of Language* (pp. 458-484). Blackwell.
- Soames, S. (1999). Vagueness, Partiality and the Sorites Paradox. In: *Understanding Truth* (pp. 203-227). Oxford University Press.
- Tye, M. (1994). Sorites Paradoxes and the Semantics of Vagueness. *Philosophical Perspectives*, 8 (Logic and Language), 189-206.
- Unger, P. (1979). There Are No Ordinary Things. *Synthese* (41), 117-154.
- Williamson, T. (1997). Imagination, Stipulation and Vagueness. *Philosophical Issues*, 8 (Truth), 215-228.

- Williamson, T. (1994). *Vagueness*. Routledge.
- Williamson, T. (1992). Vagueness and Ignorance. *Proceedings of the Aristotelian Society , Supplementary Volumes* (66), 145-177.
- Wright, C. (1976). Language-Mastery and the Sorites Paradox. In: G. Evans, & J. McDowell, *Truth and meaning: Essays in Semantics* (pp. 223-247). Oxford University Press.
- Wright, C. (1975). On The Coherence of Vague Predicates. *Synthese* (30), 325-365.
- Wright, C. (1994). The Epistemic Conception of Vagueness. *The Southern Journal of Philosophy*, XXXIII (Supplement), 133-159.
- Wright, C. (2009). The Illusion of Higher-Order Vagueness. In: D. Richard, & S. Moruzzi, *Cuts and Clouds* (pp. 523-549). Oxford University Press.

