

ESCRITOS DE FILOSOFIA IV

# Linguagem e Cognição

Marcus José Alves de Souza  
Maxwell Morais de Lima Filho  
(Orgs.)





Historicamente, a mobilização do pensamento neste âmbito de investigação filosófica, expressa neste livro, iniciou-se em 2014, com a realização do 1º Encontro Linguagem e Cognição da UFAL, que gerou o primeiro volume da série de livros de coletâneas produzidos por participantes do Encontro e por convidados: *Escritos de Filosofia: Linguagem e Cognição*, em 2015. Este processo de promover os Encontros, editar e publicar textos de colegas pesquisadores envolvidos, direta e indiretamente com o âmbito de investigação e/ou com os Encontros, repetiu-se duas vezes até o momento. Com a publicação, em 2017, do *Escritos de Filosofia II: Linguagem e Cognição* (volume 2) e, em 2019, do *Escritos de Filosofia III: Linguagem e Cognição* (volume 3). Nesta coletânea *Escritos de Filosofia IV: Linguagem e Cognição*, contamos com a contribuição de 22 pesquisadores, oriundos de 15 instituições de ensino superior brasileiras e uma estrangeira, produzindo 19 capítulos.



**FAPEAL**

FUNDAÇÃO DE AMPARO À PESQUISA  
DO ESTADO DE ALAGOAS



editora **phi**.org



## **Escritos de Filosofia IV**



# Escritos de Filosofia IV

Linguagem e Cognição

**Organizadores:**

Marcus José Alves de Souza  
Maxwell Moraes de Lima Filho



**Diagramação:** Marcelo A. S. Alves

**Capa:** Carole Kümmecke - <https://www.conceptualeditora.com/>

**Arte de Capa:** Pedro Lucena

**Revisão ortográfica e das regras da ABNT:** Marcus José Alves de Souza & Maxwell Morais de Lima Filho

**O padrão ortográfico e o sistema de citações e referências bibliográficas são prerrogativas de cada autor. Da mesma forma, o conteúdo de cada capítulo é de inteira e exclusiva responsabilidade de seu respectivo autor.**



Todos os livros publicados pela Editora Fi estão sob os direitos da [Creative Commons 4.0](https://creativecommons.org/licenses/by/4.0/deed.pt_BR) [https://creativecommons.org/licenses/by/4.0/deed.pt\\_BR](https://creativecommons.org/licenses/by/4.0/deed.pt_BR)



Dados Internacionais de Catalogação na Publicação (CIP)

---

SOUZA, Marcus José Alves de; LIMA FILHO, Maxwell Morais de (Orgs.)

Escritos de Filosofia IV: Linguagem e Cognição [recurso eletrônico] / Marcus José Alves de Souza; Maxwell Morais de Lima Filho (Orgs.) -- Porto Alegre, RS: Editora Fi, 2020.

413 p.

ISBN - 978-65-5917-030-2

DOI - 10.22350/9786559170302

Disponível em: <http://www.editorafi.org>

1. Filosofia; 2. Cognição; 3. Linguagem; 4. Lógica; 5. Mente; I. Souza, Marcus José Alves, org. II. Lima Filho, Maxwell Morais de, org.

---

CDD: 100

Índices para catálogo sistemático:

1. Filosofia                      100

*Para Bia, Jorginho,  
Tissi Camu e o Ecosistêmico Rudá*



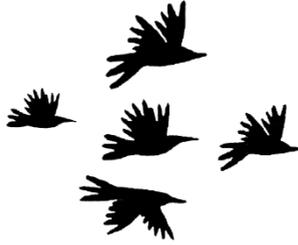
# Sumário

<b>Apresentação .....</b>	<b>13</b>
Marcus José Alves de Souza	
<b>1 .....</b>	<b>17</b>
<b>A ideia de bem, no sol, linha e caverna (REP. VI 506B-VII 535A)</b>	
José Trindade Santos	
<b>2.....</b>	<b>44</b>
<b>Como livrar-se da lógica e ser feliz</b>	
Desidério Murcho	
<b>3.....</b>	<b>64</b>
<b>Princípio da explosão e outras esquisitices lógicas</b>	
Frank Thomas Sautter	
<b>4.....</b>	<b>79</b>
<b>Desacordos profundos e revisão da lógica: uma proposta neo-pragmatista</b>	
Marcos Silva	
<b>5.....</b>	<b>100</b>
<b>A teoria supervalorativista da vagueza e o problema da precisão</b>	
Sagid Salles	
<b>6.....</b>	<b>120</b>
<b>Abstração em Frege</b>	
Alessandro Bandeira Duarte	
<b>7.....</b>	<b>139</b>
<b>Neodescritivismo sobre termos gerais</b>	
Claudio Costa	

<b>8</b> .....	<b>158</b>
<b>Lewis contra Putnam, Fodor, Davidson e Kripke</b>	
Rodrigo A. dos S. Gouvea	
<b>9</b> .....	<b>186</b>
<b>Quine e a dispensabilidade da identidade</b>	
Hudson dos Anjos Benevides	
<b>10</b> .....	<b>208</b>
<b>Uma nota sobre a relação entre filosofia da mente e filosofia da linguagem</b>	
João de Fernandes Teixeira	
<b>11</b> .....	<b>217</b>
<b>Existem sensações inconscientes? Um exame da teoria das sensações de David Rosenthal</b>	
Tárik de Athayde Prata	
<b>12</b> .....	<b>241</b>
<b>Representacionalismo e realismo direto na teoria da percepção de John R. Searle</b>	
João Paulo M. de Araujo	
<b>13</b> .....	<b>266</b>
<b>Relações causais e racionais entre mente e mundo: a resposta de Donald Davidson a John McDowell</b>	
Marco Aurélio Sousa Alves	
<b>14</b> .....	<b>292</b>
<b>O empirismo cognitivo de Sylvain Auroux</b>	
Argus Romero Abreu de Moraes	
<b>15</b> .....	<b>316</b>
<b>Aproximações entre direito e moral nos conflitos entre normas</b>	
Juliele Maria Sievers	

<b>16</b> .....	<b>334</b>
<b>Explicações psicológicas em psiquiatria: o caso da neuropsiquiatria cognitiva</b>	
Rondineli Bezerra Mariano	
Richard Theisen Simanke	
<b>17</b> .....	<b>362</b>
<b>Linguagem e psicologia em Ludwig Wittgenstein</b>	
Thiago Ferreira dos Santos	
<b>18</b> .....	<b>377</b>
<b>A descrição dos próprios estados mentais</b>	
Christopher Alexander Franke	
Joelma Marques de Carvalho	
<b>19</b> .....	<b>399</b>
<b>Pensamento “crítico” &amp; pseudo-história</b>	
Ricardo Seara Rabenschlag	
<b>Sobre os Organizadores</b> .....	<b>413</b>





## Apresentação

*Marcus José Alves de Souza*

Relações entre estes conceitos formam uma paisagem  
com que a linguagem nos presenteia  
em inúmeros fragmentos;  
juntá-los é-me *demasiado difícil*.  
Consigo apenas fazer um trabalho muito imperfeito!

**L. Wittgenstein**

Apresentar um livro coletânea de Filosofia como este é, no fundo, apresentar um âmbito de investigação, com vários autores que conversam com autores, através de conceitos e argumentos relacionados a um conjunto de problemas. Como está escrito na epígrafe desta apresentação, articular os fragmentos que a linguagem nos presenteia é uma tarefa difícil, feita imperfeitamente; no caso da tarefa de apresentar este livro como um todo, tenho apenas fragmentos, poucas palavras que, imperfeitamente, falam deste âmbito de investigação na Filosofia: *significado, mente, lógica, consciência, linguagem, conhecimento, ciência...* Tais fragmentos de linguagem, neste contexto, significam também os múltiplos desdobramentos reflexivos e problemas que estas palavras guiam.

Talvez ter um âmbito de investigação pelas palavras que indiquei seja a maior virtude deste livro e da nossa série como um todo. É a maior virtude porque, com este âmbito estabelecido, pode-se mobilizar o pensamento a avançar no aperfeiçoamento da delimitação dos problemas

que envolvem estas palavras na Filosofia ou mesmo avançar a investigação lançando problemas novos, novas relações ou mesmo ainda dissolvendo problemas antigos.

Historicamente, a mobilização do pensamento neste âmbito de investigação filosófica, expressa neste livro, iniciou-se em 2014, com a realização do 1º Encontro Linguagem e Cognição da UFAL, que gerou o primeiro volume da série de livros de coletâneas produzidos por participantes do Encontro e por convidados: *Escritos de Filosofia: Linguagem e Cognição*, em 2015. Este processo de promover os Encontros, editar e publicar textos de colegas pesquisadores envolvidos, direta e indiretamente com o âmbito de investigação e/ou com os Encontros, repetiu-se duas vezes até o momento. Com a publicação, em 2017, do *Escritos de Filosofia II: Linguagem e Cognição* (volume 2) e, em 2019, do *Escritos de Filosofia III: Linguagem e Cognição* (volume 3). Ao todo, nestes três volumes, tivemos a contribuição de 40 pesquisadores, produzindo 37 capítulos, provenientes de 12 instituições de ensino superior do Brasil, especialmente, do Nordeste brasileiro.

O mesmo processo se repetiu mais uma vez para a confecção desta coletânea, *Escritos de Filosofia IV: Linguagem e Cognição*, que ora apresentamos. Nela, contamos com a contribuição de 22 pesquisadores, oriundos de 19 instituições de ensino superior brasileiras e uma estrangeira, produzindo 21 capítulos. Neste quarto volume, o leitor poderá, inicialmente, retomar uma reflexão apurada dos significados de elementos da alegoria da caverna. Seguirá refletindo sobre temáticas da lógica, seja retomando a problemática conceitual da validade lógica, seja tratando criticamente do princípio de explosão, seja avaliando os desacordos profundos por um viés pragmático, além de, mais uma vez, apreciar os limites da vagueza (vacuidade) e da precisão conceituais em lógica. O leitor terá a oportunidade de visitar e revisar temas e autores tradicionais da Filosofia Analítica, como o tema da explicação (neo)descritivista de termos gerais; elementos básicos do pensamento de alguns autores como Frege, Quine e Lewis, além de Putnam, Fodor, Davidson e Kripke. Na sequência,

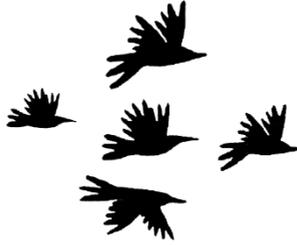
poderá repensar as interfaces entre Filosofia da Linguagem e da Mente; poderá ter elementos conceituais interessantes para avaliar a questão do inconsciente, mais especificamente, a questão das sensações inconscientes, a partir do filósofo estadunidense David Rosenthal; poderá discutir sobre a temática do representacionismo no pensamento de John Searle; e sobre as relações entre mente e mundo na contraposição de dois grandes pensadores: Donald Davidson e Jonh McDowell. Por fim, poderá ainda questionar-se sobre as relações entre moral e direito, a partir de três posições normativas do direito: positivismo, naturalismo e “justiça poética”, esta última menos conhecida, mas que coloca o conceito de empatia como central na resolução dos conflitos normativos. Poderá conhecer a proposta de empirismo cognitivo de Sylvain Auroux. Do mesmo modo, refletir sobre a instigante relação entre os conceitos filosóficos sobre o mental e a construção teórica da psiquiatria. Avaliar aspectos proposta de Wittgenstein sobre dimensões do mental: seja de um modo mais geral, seja discutindo um aspecto específico da descrição dos próprios estados mentais. E por fim, apreciar uma reflexão acerca da problemática do processo histórico-filosófico de construção da ciência moderna, através de uma crítica a determinadas concepções (“mitos”) sobre este processo. Nesta breve panorâmica, percebe-se o valor da diversidade de temáticas, de autores, de abordagens e da interligação de áreas de investigação que este volume contempla. É uma pequena amostra da vitalidade do pensamento filosófico em torno do “âmbito” que comentávamos no início desta apresentação, assim como é expressão da força de investigação de nossos colegas pesquisadores de Filosofia.

Além dos 4 volumes desta Série, os organizadores também editaram o Dossiê Linguagem e Cognição, na Revista *Prometheus* (<https://www.seer.ufs.br/index.php/prometeus/issue/view/861>), em 2020, o que mostra capacidade de mobilização de pesquisadores, perseverança editorial e temática nas publicações. Os organizadores também lideram o Grupo de Pesquisa Linguagem e Cognição da UFAL, como foco de fomento à investigação dessas temáticas, além de serem os

idealizadores e promotores, junto com os demais membro do Grupo, dos vários encontros realizados até o momento: já são sete edições.

Agradecemos aos autores que forneceram os textos para confecção deste volume. Temos muito prazer de termos colegas de reconhecimento nacional e internacional prestigiando e compondo esta coletânea, como os professores: José Trindade dos Santos, Claudio Costa, Desidério Murcho e João de Fernandes Teixeira. Agradecemos ao apoio institucional do Instituto de Ciências Humanas, Comunicação e Artes, do Curso de Filosofia da UFAL (Licenciatura e Mestrado), a todos(as) que contribuem com as atividades do Grupo de Pesquisa e à Editora Fi pela parceria nas publicações.

Maceió, novembro de 2020.



**A ideia de bem, no sol, linha e caverna**  
**(*REP.* VI 506B-VII 535A)**

*José Trindade Santos*<sup>1</sup>

Embora a “Ideia do bem” seja introduzida no debate político, na *República* VI-VII, como objeto de estudo (*megiston mathêma*: R. VI 504e7-505b3), a atenção que lhe será dispensada entre 504e e 535a aponta à natureza transcendente que a situa para lá da realidade e do saber. Cremos que essa condição justifica as reticências de Sócrates em expor – nas palavras de Gláucon – “uma [doutrina que é] dele” (506b6-7; 506b-e). O filósofo prefere recorrer a analogias para chegar ao bem, recorrendo a uma linguagem densamente povoada de metáforas.

Por outro lado, da perspectiva da composição dialógica, é sabido que Platão usa a analogia como estratégia adequada à investigação do inteligível, uma vez que a natureza invisível deste obriga a usar o pensamento para o atingir. Opta então por recorrer à experiência que os interlocutores têm da exploração do visível para projetar no inteligível informações daquele retiradas. Será, pois, com essa finalidade, que esboça

---

<sup>1</sup> Professor Catedrático da Universidade de Lisboa (ULisboa) e Professor Visitante da Universidade Federal do Ceará (UFC), de cujo Programa de Pós-Graduação em Filosofia é Professor Permanente. Tem experiência na área de Letras, com ênfase em Filosofia Grega Antiga, atuando principalmente nos seguintes temas: Epistemologia, Filosofia da Linguagem e Metafísica. Orcid: <https://orcid.org/0000-0002-0631-4348>

um extenso e diversificado argumento, em três fases, distribuído por duas analogias e uma alegoria.

Na primeira, parte de um exemplo bem conhecido: os efeitos gerados pela luz do Sol no mundo visível, que metaforicamente transfere para a cognição. Na segunda, recorre à relação original/cópia, uma vez mais usando a análise do visível para dele projetar o acesso ao inteligível. A finalizar, usa a alegoria, que descreve a existência de escravos aprisionados numa Caverna, para contrastar duas atitudes sobre a vida e o conhecimento, nos âmbitos psíquico e político.

### A “analogia do sol”

Na “analogia do Sol” (508a-509b), a síntese do argumento de Sócrates é atingida quando a Ideia de bem (ou “Forma do Bom”) é apresentada como causa dos inteligíveis e da sua possibilidade de serem conhecidos, conferindo unidade e finalidade ao sistema estruturado das Ideias.

No entanto, o filósofo já tinha iniciado a exploração do processo cognitivo no argumento final do Livro V, onde argumenta a favor das Ideias, ao mostrar a um interlocutor imaginário que as opiniões visam às “coisas que são e não são” (476a-479c; em particular, 478d-e; Annas, 1981, p. 190-216; Harte, 2011, p. 202-6). Pois, pelo fato de participarem de Ideias contrárias (justo/injusto, belo/feio, etc.), pressupõem a existência de Ideias, “sempre permanentemente em si” (479a1-3; *Phd.* 78d6-8; Santos, 2020).

Volta aqui a este ponto, contrastando as multiplicidades visíveis, distintas pelo discurso, com a unidade de cada uma das Ideias “que é” (*Phd.* 102b1; *R.* X 596a5<sup>2</sup>), cuja natureza inteligível a torna apenas acessível ao pensamento (507b-c). Mediante exemplos da captação de audíveis e visíveis introduz então a noção da “competência perceptiva” (*dynamin*:

---

<sup>2</sup> A “distinção [de cada multiplicidade] pelo discurso” refere a atribuição a cada uma do nome da respectiva Ideia: “Para cada multiplicidade à qual aplicamos um mesmo nome costumamos postular uma certa ideia e só uma” (596a5-7).

507c10; R. V 477c ss.) que lhes corresponde. Mas o seu objetivo é chegar mais acima.

Tendo então distinguido a percepção da competência que permite a sua captação, alude a uma “terceira espécie” – sem a qual aquela não é possível (507d-e) –, de modo a, se concentrando na visão, chegar à luz, e através dela ao Sol (508a), identificado como o “responsável pela vista” (508b). Até que, tendo-o caracterizado como “filho do bem”, pode, a partir do visível, visar ao seu análogo inteligível (508b-c):

Fica sabendo que o que transmite a verdade aos objetos cognoscíveis e dá ao sujeito que conhece esse poder é a ideia do bem. Entende que é ela a causa do saber e da verdade, na medida em que esta é conhecida, mas, sendo ambos assim belos, o saber e a verdade, terás razão em pensar que há algo de mais belo ainda do que eles. E, tal como se pode pensar correctamente que neste mundo a luz e a vista são semelhantes ao Sol, mas já não é certo tomá-las pelo Sol, da mesma maneira, no outro, é correcto considerar o saber e a verdade, ambos, semelhantes ao bem, mas não está certo tomá-los, a um ou a outra, pelo bem, mas sim formar um conceito ainda mais elevado do que seja o bem<sup>3</sup> (508e-509a).

E conclui:

Logo, para os objectos do conhecimento, dirás que não só a possibilidade de serem conhecidos lhes é proporcionada pelo bem, como também é por ele que o Ser e a essência lhes são adicionados, apesar de o bem não ser uma essência, mas estar acima e para além da essência, pela sua dignidade e poder (509b).

Estes dois excertos do Livro VI têm, desde a Antiguidade, sido objeto de acérrimas disputas hermenêuticas e doutrinárias, pelo que requerem cuidadosa análise. Começamos pelo primeiro.

A analogia é bem sugestiva. Se a luz é a causa de algo ser visto, e o Sol – análogo do Bem – a fonte da luz, então, no inteligível, a Ideia do bem é “a causa do saber e da verdade”. A ascensão, através dos olhos (análogos à alma) e da visão (à inteligência) à luz (à verdade) (508c5, d5),

---

<sup>3</sup> Recorremos à tradução de M. H. da Rocha Pereira, indicada nas Referências, com pequenas modificações.

permite o acesso à transcendência do saber. Mas não se detém aí, na fronteira que a separa do visível. Pois, tal como a luz e a vista se assemelham ao Sol, mas não se confundem com ele, assim também saber e verdade são semelhantes ao bem, mas não podem ser tomados por aquele que em dignidade os excede (508e-509a).

Além dos termos análogos (acima esclarecidos nos parênteses), o excerto é pesadamente carregado de metáforas, algumas das quais terão de ser decifradas sob pena de inviabilizarem a compreensão do passo, pese embora o risco de erro da interpretação. A mais evidente será ‘verdade’. Pelo que o passo sugere, o termo formula, em dois sentidos complementares, a condição que permite “conferir a possibilidade de conhecer” (ao cognoscente) e garantir a “autenticidade” (dos cognoscíveis<sup>4</sup>).

Relevante também será o termo ‘causa’. Adiante, numa pontual recapitulação da primeira analogia, Sócrates esclarece que a “ideia do bem é a última coisa a ser vista, ... causa de tudo o que é justo e belo...”, “nos [mundos] visível e inteligível”, no sentido de os ter criado e de deles “ser senhora” (VII 517b8-c2). Todavia, indissociável destes dois, outro sentido do termo será o de “explicação”: aquilo que, no coração da analogia, permite que “aquilo que é”, como tal, seja acessível à cognição (Vlastos, 1981).

Esta superioridade da causa, em relação àquilo de que é causa, obriga, portanto, a “formar um conceito ainda mais elevado do que seja o bem” (509a4-5). É para possibilitar o esclarecimento desta exigência que, após ter notado que, além de permitir a vista das coisas, o Sol ainda lhes confere “gênese, crescimento e alimentação” (509b), é apresentado o segundo excerto (509b).

Começando por reafirmar o princípio exposto de que é pelo bem que a cognoscibilidade é proporcionada ao “sujeito” e aos “objetos” do

---

<sup>4</sup> Em ambos os casos, a concessão é condicional, atuando nos dois planos cognitivos: a verdade é concedida aos “que podem conhecer”, de acordo com a “competência cognitiva” à qual recorrem, e às [coisas] “que podem ser conhecidas”. De um ponto de vista epistemológico, esta ‘verdade’ parece se identificar com “o que é”, seguindo Parmênides (B8.18b).

conhecimento (508e1-2), o passo acrescenta que, além desta função, o bem ainda lhes confere “ser e essência”, embora ele próprio se ache “para além da essência”. E é aqui que os comentadores se digladiam pois os termos e a expressão usados consentem uma ampla diversidade de interpretações.

É claro, já desde o final do Livro V (476e *ad fin.*), que a concepção de conhecimento aí proposta é marcada pela dualidade, e de múltiplos pontos de vista. Pois, não só relaciona um sujeito e um objeto (*ti*) cognitivos (476e7), como atua nos distintos registros das competências ontoepistemológicas<sup>5</sup> designadas como “opinião” e “saber”. De resto, o argumento enfaticamente insiste no fato de essas duas competências cognitivas se distinguirem uma da outra, em diversos aspectos.

Embora a informação não seja explicitamente avançada no argumento, é indisputável que, no caso da “opinião”, o lugar do ‘sujeito’ cabe àquele que exerce a competência opinativa (associada à perceptiva: 477c3-4), e àquele “que sabe”, no caso do “saber”. Por outro lado, no que respeita ao objeto da cognição, o texto distingue nitidamente as duas competências mediante dois índices: “aquilo sobre que se exercem” e os “efeitos que produzem”; a saber: num caso, a “aparência” (*doxa*) e as “opiniões” (*doxai*); no outro, “o que é” e o “saber”. Portanto, no passo do Livro VI que estamos a analisar, é claro que aquilo que é captado pelos olhos, no mundo “visível”, por ser iluminado pelo Sol, são “aparências” afins às “opiniões”, enquanto que as Ideias inteligíveis, captadas pela inteligência, por ação da Ideia do bem, serão “o que é”, captado pelo “saber”.

Restam ainda três dificuldades: os termos ‘ser’ e ‘essência’ e a expressão “para além da essência”. No caso dos termos, não cremos que adiante muito recorrermos ao Grego para sermos esclarecidos, pois, o fato de terem recebido diferentes traduções, em diversas línguas, contribuiu para que entre tradutores e comentadores não houvesse consenso sobre a sua interpretação. Teremos, portanto, de conjecturar o que Platão terá

---

<sup>5</sup> Pois não só se referem a realidades existentes, como às competências habilitadas a captá-las.

querido dizer com “o ser” e a “essência”. Começamos pelo primeiro, para dele passar ao outro.

A opção pela tradução ‘o ser’ (*to einai*: e não *to on*, note-se) é evidentemente correta. No entanto, aqui a dificuldade não está na escolha da expressão adequada, mas no que Platão terá querido dizer com ela, neste contexto. Aceitando o argumento do Livro V, o “ser” de uma coisa aponta para “o que ela é”. No caso de uma Ideia (ou Forma), na *República* (como no *Fédon*) trata-se sobretudo de predicados (G. Owen, 1957, 103-11; embora, noutros contextos, possa também se tratar de naturezas: por exemplo, *R. X* 596a5-7; *Parm.* 130c). Prestaremos atenção apenas à primeira possibilidade.

Afirmar que uma Ideia é um predicado, nos planos sintático e semântico, levanta sérios problemas lógicos, pois, se, em si, a Ideia é uma “entidade” — algo “que é” —, portanto, acima de tudo um ‘sujeito’, como é que pode *também* ser um ‘predicado’? A resposta platônica (ao problema posto por gerações de aristotélicos) é que isso é possível porque a Ideia é também uma “essência” (*ousia*); ou seja, algo que não apenas “é”, como também “é aquilo que é” (*auto hekaston ho estin: Phd.* 78d4). Significa isto que a Ideia é uma entidade que se autopredica<sup>6</sup>; ou seja, que, quando se sustenta que a Ideia é *um predicado*, se diz que ela é aquilo que, além de se dizer de muitas outras “coisas” (que a *têm* como *predicado*), se diz também dela própria<sup>7</sup>.

Estes esclarecimentos devem ter contribuído para a compreensão dos sentidos dos termos usados. No entanto, é oportuno mencionar outras traduções correntes para *to einai* — “existência” (E. Chambry, 1933; P. Shorey, 1935; ca. Vegetti, 2003: “essere” e “essenza”) — e “realidade” (Shorey, 1935, traduz *to on* por “reality”: 508d5).

---

<sup>6</sup> Não confundir ‘predicação’ com ‘autopredicação’.

<sup>7</sup> Por exemplo, “a justiça é justa” é um caso de autopredicação, implicando que “a justiça” é a única entidade à qual cabe irrestritamente o predicado “justa”. Pelo contrário, qualquer “ação ou corpo” não será “justa” — apesar de participar da justiça — pelo fato de também participar da injustiça (*R. V* 476 a ss.).

Esta última tradução atende ao contexto metafórico em que as competências cognitivas são avaliadas, segundo a sua “clareza” ou “obscuridade” (*R. V 478c, VI 507e, 509d-e, passim*), sendo atribuída maior ou menor realidade àquilo que se percebe de modo mais ou menos claro (Vlastos, 1981a). A primeira força um pouco a terminologia platônica pelo fato de isolar e separar a leitura existencial de *einai* de outras, pressupondo que aquilo que “de todas as maneiras é” (*pantelôs on: 477a, passim*), em qualquer leitura do verbo, não pode deixar de “existir”.

Resta a maior dificuldade, a da expressão “para lá da essência”. Se aceitarmos que uma essência pode ser vista como um predicado monádico que se diz de si próprio, teremos de admitir que “para lá” dele ficará a sua “essencialidade”; ou seja, o conjunto de propriedades que as essências possuem, enquanto essências, e é comum a todas elas. Poderemos referir a unidade, imutabilidade, perfeição etc. — as chamadas “propriedades eidéticas” —, cujo *locus classicus* é o passo do *Banquete* 211a-c; ver: *Phd.* 78c-79a). Portanto, a tese sustentada no excerto em apreço é a de que essas propriedades se deverão à participação de todas as Ideias na estrutura teleológica proporcionada pela participação das Ideias no bem (Fine, 1999, p. 228-9, p. 244; G. X. Santas, 1999, p. 249-68; N. Denyer, 2007, p. 284-92). Princípio que, por não ser uma essência, não poderá ser incluído no conjunto de todas as Ideias que dele participam; nada impedindo, porém, que possa ser conhecido por meio de enunciados (533c-535a<sup>8</sup>; White, 1979, p. 181, p. 200-1). Esta interpretação será aceitável por quantos defendem que a concepção platônica de saber é proposicional (por exemplo, J. Annas, 1981, p. 242-3 — “explanatória”, por envolver “compreensão” —; Fine, 1999, p. 217-46; explicitamente: p. 225); sustentando que terá de ser pela apresentação de proposições (*logoi*) que alguém pode mostrar “que sabe” (*R. VII 534b-c, passim*).

---

<sup>8</sup> Notar a insistência explícita na necessidade de “dar razões” (*logon didonai*: 531d6, 533c4-5, 534b4-5), de “definir pelo enunciado” (*diorisasthai tōi logōi*: 534b9) a Ideia de bem; de captar pelo enunciado (*haptōti tōi logōi*: 534c3) o saber do bem, por mais que Gláucôn considere essa tarefa um “labor imenso”: 531d6.

Mas há muitos intérpretes — por exemplo, médio e neoplatônicos —, que defendem uma concepção em que o saber é diretamente captado por contato (Plotino, *Enn.* V,3,10), para quem a expressão usada implica que Platão terá, ao menos, aberto a porta à tese de que há algo “para além do ser”, inacessível à racionalidade discursiva. Nada há, no texto, que certifique esta conclusão. No entanto, no *Parmênides*, com a primeira hipótese sobre o uno (137c-142a), o filósofo terá viabilizado essa suspeita, ao mostrar que, separado da essência, “que é”, do um nada pode se dizer, senão “se um um” (142c3). Pois, só “se um é”, participando da essência (Hipótese II: *hen ei estin* — 142c8), será possível lhe atribuir todos os predicados negados na Hipótese I.

Ora, será a aceitação desta linha de argumentação que irá permitir a Plotino, por um lado, sustentar a anterioridade do uno ao ser, e, por outro, defender que “nada poderia ser, se não fosse um” (*Enn.* VI.9.1,3 ss.; Platão *Sofista* 245d). A partir da conjunção destas duas teses, consegue então esboçar uma via de acesso ao um, mostrando como, por si próprio, o intelecto (*nous*) é capaz contemplar a natureza do um, suspendendo (à semelhança de Platão, na Hipótese I, do *Parmênides*) todos os predicados atribuíveis ao ser (*Enn.* VI.9.1.3-4), se tornando no um quando pensa (V.3.11,13-15).

Retornam por isso as perguntas sobre o sentido que deve ser atribuído à “Ideia do bem”. Por que razão Sócrates recorrerá a metáforas superlativas para falar dele (“mais brilhante”, “mais feliz”, “mais excelente”: VII 518c9, 526e3-4, 532c5-6)? Porquê então, no momento em que Gláucon lhe pede que não abandone a pesquisa, se ter escusado a responder “o que é o bem em si”<sup>9</sup> (506d-e)? Como se explicam estas reticências, se, no termo do percurso que o conduziu à dialética, tanto insiste sobre a necessidade de conhecer e definir o bem<sup>10</sup>?

---

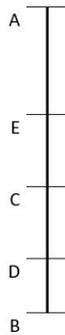
<sup>9</sup> Encontrando no passo — 504e7-505b3 — apoio para entender a Ideia de bem como “princípio da realidade”, consubstanciando o universal “desejo do bem”, El Murr (2014, p. 58-9) justifica a atitude de Sócrates perante o desafio posto pelo *megista mathêma*.

<sup>10</sup> G. Fine (1999, 226-227) observa que a escusa de Sócrates se deve a ter opiniões, mas não o saber do bem (506c), cuja função epistemológica e metafísica passará de seguida a explicar. Mais tarde, “poderá dispensar as imagens” (510b).

Estas perguntas são deixadas no ar, habilitando qualquer interpretação. No entanto, entre as duas opções mais favorecidas pelos comentadores — a esotérica<sup>11</sup> e a que alega a dificuldade da problemática<sup>12</sup> — preferimos a segunda.

### “Analogia da linha” I

A “Linha dividida” é oferecida para esclarecer dúvidas que a analogia do Sol tenha deixado a Gláucon (509c). Uma linha cortada em duas partes desiguais (509e) representa os dois mundos — da realidade e da cognição —, nos quais o Sol e a Ideia do bem reinam. Subdivide-se então cada uma das seções em dois segmentos (509e-510a):



A linha vertical (A-B) condensa a realidade e o conhecimento que dela há. As duas seções (A-C, C-B) representam os mundos inteligível e visível, e os quatro segmentos associam-nos às modalidades cognitivas pelas quais podem ser captados, comparando originais com imagens. Obtêm-se assim quatro segmentos verticalmente dispostos, de cima para baixo:

<sup>11</sup> Por exemplo, T. A. Szlezák (1985); G. Reale (1997).

<sup>12</sup> Por exemplo, G. Fine, 1999; G. X. Santos, 1999; M. Vegetti, 1993; W. Wieland, 1982, p. 159-85; W. Guthrie, 1975, p. 507.

A—D

D—C

C—E

E—B

De baixo para cima, ‘E—B’ representa as imagens do mundo visível: sombras, imagens (*phantasmata*: 510a2) nas águas, em espelhos. ‘C—E’ representa o “nosso mundo”, do qual o anterior é imagem (509e-510a). Esta identificação permite esboçar a analogia. Em conjunto, os dois segmentos representam o mundo visível, e a relação entre ‘C—E’ e ‘E—B’ contrapõe “o que é verdadeiro” (“os originais”<sup>13</sup>) a “o que não é verdadeiro” (as “imagens”: *eikones*). Anunciando a passagem ao inteligível, à guisa de síntese da análise do visível, para selar esta contraposição, Sócrates remata: “tal como o opinável está para o cognoscível (*doxaston/gnôston*: ver 476a-478e), assim está “o que se assemelha” em relação àquilo “a que se assemelha” (510a11).

Aceite-se então análoga relação entre ‘D—C’ e ‘A—D’, no inteligível. No segmento inferior, usando “como imagens” os [originais], atrás (em ‘C—E’) imitados (em ‘E—B’), “a alma”:

- a) investiga a partir de hipóteses;
- b) não caminha para o princípio, mas para a conclusão;
- c) recorre a imagens.

Pelo contrário, no segmento superior, a alma:

- a) “parte da hipótese para chegar ao ‘princípio não-hipotético’”;
- b) “dispensa imagens”, avançando “com o auxílio das ideias” (510b).

O passo comprime tantas informações importantes que Gláucon abertamente manifesta a sua incompreensão. Sócrates decide então

---

<sup>13</sup> O termo grego que, seguindo procedimento corrente em tradutores e comentadores, traduzimos como ‘originais’ não aparece no passo citado (510a10-11), são aí comparados “o que se assemelha” a “aquilo a que se assemelha”.

auxiliá-lo, propondo como exemplo a prática dos geómetras (para o que é útil relembrar o diálogo com o escravo, no *Mênon* 82-86, a que o texto indiretamente alude). No *Mênon*, o quadrado, a regra da sua área, a diagonal que o divide em dois triângulos retângulos iguais, são hipóteses que o escravo acolhe sem justificação. Usando as figuras desenhadas, Sócrates esboça raciocínios que o moço é levado a aceitar. Nesse ponto, porém, se afastam um do outro. Pois, enquanto o rapaz opera exclusivamente sobre a figura, se apoiando nela para responder ao filósofo, este raciocina sobre “o quadrado e a diagonal, em si” e não sobre os desenhados (*R.* 510d-e), se servindo das imagens para “ver o que não pode avistar-se, senão pelo pensamento” (510e).

O exemplo mostra como os dois “se servem de hipóteses”... “sem ir ao princípio”, recorrendo a imagens, porque, “em comparação com as sombras” (“os objetos imitados na seção inferior”: E-B), ...”estas são mais claras” (511a). No entanto, como se servem de hipóteses não-explicadas, “não conseguem ir ao princípio”, ou seja, atingir o saber sobre o quadrado (ou responder à pergunta: “o que é a virtude?”), se ficando pela conclusão (que responde ao problema “da linha a partir da qual se constrói o quadrado de área dupla” apontando a diagonal).

## Síntese

Da nossa perspectiva, a ontoepistemologia proposta pelas duas analogias atinge aqui o seu ponto mais elevado, pois o tratamento conferido ao segmento superior da seção do inteligível nos parece meramente projetivo, anunciando a introdução da dialética. De resto, é aqui que a Linha inova em comparação com o Sol. Como dissemos atrás, ambas as analogias exploram o visível para ter acesso ao inteligível. Todavia, a Linha não só leva a cabo uma análise do visível — no Sol meramente introdutória —, como aponta a possibilidade de trânsito entre os dois mundos.

Se percebe que a Linha começa a se afastar do Sol ainda na análise do visível, quando Sócrates inopinadamente reintroduz o par de termos relevante no argumento que termina o Livro V — “opinável” e “cognoscível” —, para referir “aquilo sobre que se exercem” as competências cognitivas “opinião” e “saber”. O passo nos parece de importância capital para operar a inserção das analogias dos Livros VI e VII no contexto do argumento que formalmente investe a opinião com capacidade cognitiva (tomando o lugar ocupado antes pelo *logos* e superando a alternativa ‘aporia’/‘infalibilidade’ — típica da obra “socrática” — pela contraposição da verdade à falsidade: VI 510a9).

A revolução epistemológica que a ele se segue acha-se condensada no passo 510b, incompreensível na forma comprimida com que o texto o oferece ao leitor. Nesse breve parágrafo, confrontam-se os dois segmentos abarcados pela seção do inteligível. Mas a função dinamizadora da cognição, comandada pela *dianoia*, passa despercebida nas duas linhas de Grego que se referem à seção anterior (510b4-6).

A informação mais importante diz respeito ao fato de os raciocínios dependerem de hipóteses e do *fim* por eles visado — a conclusão [que buscam] —, em oposição ao “princípio não-hipotético” (o bem), visado pelo segmento superior (510b6-8). No entanto, o dinamismo que caracteriza esta “afeção na alma” (511d6) é explícito na sua capacidade de “usar as coisas *então imitadas* [no segmento inferior do visível, supõe-se] como se fossem imagens [das Ideias]”.

Significa isto que a *dianoia* transforma em imagens (não necessariamente apenas visuais) os “originais” imitados pelo segmento inferior do *sensível* (já não apenas do visível: por isso Sócrates lhe chamou “opinável”). E para quê o faz? Decerto para usar essas imagens para ascender ao inteligível, buscando as Ideias que, por sua vez, as imagens inteligíveis imitam. Todavia, só captamos a compreensão da amplitude do movimento reflexivo no passo se aceitarmos a sugestão nele implícita de recorrermos a imagens visíveis das Ideias para ascender ao inteligível, como faz Sócrates no diálogo com o escravo, no *Mênon*.

Parece-nos que a relevância deste texto dificilmente poderá ser exagerada, uma vez que contém em embrião toda uma teoria do conhecimento, ao projetar no inteligível, ou seja, tornar acessível ao pensamento toda a espécie de dados colhidos do *sensível*, agregando às imagens visuais a variedade de imagens colhidas de outros sentidos, nomeadamente do ouvido, juntamente com os juízos expressos nas opiniões produzidas pelo discurso.

É importante que nos demos conta do dinamismo do recurso à ‘imitação’, o qual, implicando uma diversidade de operações cognitivas, viabiliza a continuidade da ascensão, do nível mais baixo do sensível ao inteligível, condensada nas imagens que *intermediam* sensível e inteligível (esquemas geométricos e números). Mas há que não esquecer que, embora o acesso ao inteligível refira um movimento ascendente, como o Sol deixou claro, o impulso que dinamiza o sistema age de cima para baixo (da Ideia do bem até às “sombras” visíveis).

Põe-se então a pergunta: como é possível aceder às Ideias insusceptíveis de representação visível? Só após a crítica à Caverna, encontraremos uma tentativa de resposta a esta pergunta. Mas será possível, com risco de erro, supor que aqui intervém a reminiscência, como o *Fédon* e o *Fedro* advogam, pois não podemos esquecer que as Ideias “estão na [alma]”, se manifestando a quem as reconhece, quando estimuladas pela sensopercepção (*Fédon* 72e-77a). Como, porém, a *República* nenhuma referência explícita faz a esta possibilidade, devemos passar ao segmento superior do inteligível, já referido.

## **A Linha II: conclusão**

Que significa postular que a alma “parte da hipótese para chegar ao “princípio não-hipotético” e “dispensa imagens”, avançando “com o auxílio das ideias” (510b)? Terminado o exemplo dos geómetras (510b-511b), o texto retoma aí o segmento superior do inteligível. A segunda especificação não levanta dificuldades. Implica apenas que os raciocínios

são feitos a partir de hipóteses e não recorrem a qualquer espécie de exemplificação colhida do sensível, relacionando exclusivamente ideias.

A primeira, contudo, requiere algum esclarecimento prévio. Para começar, as hipóteses não podem simplesmente ser aceitas como evidentes, devendo ser questionadas e justificadas pela argumentação. Se for aceito o princípio do *Fédon* (100a), lhes é provisionalmente conferida verdade enquanto concordarem umas com as outras, devendo ser rejeitadas como falsidades quando o desacordo entre elas se manifestar.

Mediante a concatenação das hipóteses, deverão ser atingidas outras, cada vez “mais elevadas”, até se chegar a uma considerada “bastante” para confirmar a investigada<sup>14</sup> (*Fédon* 101c-e). Mas o término de todas as pesquisas é o “princípio não-hipotético” – se supõe que seja o bem –, cujo conhecimento só então deverá conferir verdade ao sistema hipotético anteriormente formulado, podendo daí a investigação *descer* até à conclusão buscada (511b-c; as referências a “acima/“abaixo”, “subir/“descer” caracterizam o movimento dialético; por exemplo, na *R.* 511a6, b9, *passim*; *Phd.* 101d8).

Embora ainda lhe restem dúvidas, Gláucon capta o foco do raciocínio de Sócrates. Coroando todas as pesquisas, a dialética usa o raciocínio para avançar na teia hipotética, mas a busca só atinge o seu termo quando “*sobe* até ao princípio”<sup>15</sup>. Começando pelo segmento inferior do inteligível – entre a opinião e a inteligência (*noêsis*) –, o “entendimento” (*dianoia*) é inferior à “inteligência”, mas superior à “crença” e ao “imaginário”. A hierarquização é justificada pelo grau de clareza atribuído às “afeções na alma”, de acordo com a “verdade dos seus respectivos objetos” (511d-e). O projeto da dialética só será, porém, contextualizado no processo educativo

---

<sup>14</sup> Um dos raros exemplos desta situação será o argumento final do *Fédon* sobre a imortalidade da alma. O raciocínio parte da relação entre contrários sensíveis – fogo e neve –, passa ao par e ímpar, estendendo-se daí aos números, até formular a hipótese da “imperecibilidade da alma”, considerada bastante para confirmar a inicialmente apresentada.

<sup>15</sup> Para ter uma noção sobre o acesso à Ideia de bem, bem como da função que desempenha no projeto da dialética, ver N. P. White (1976, p. 99-103, com as respetivas notas). O carácter “projetivo” desta proposta é evidente na observação de que o bem, mais do *apreendido* deve ser *reconhecido* como condição para a *identificação* de qualquer outra Ideia (p. 101-3).

depois de o filósofo ter delineado a sua proposta de um currículo de estudos visando à formação dos cidadãos (VII 523a-534d); sobretudo aos futuros guardiões e filósofos.

### **A alegoria da caverna**

A Caverna é a terceira analogia da *República*. Pelo fato de iniciar o Livro VII, aparentemente corta com a Linha e o Sol, quando, pelo contrário, as prolonga, recontextualizando-as numa perspectiva dinâmica (Raven, 1965, 131-187). Prosseguindo o projeto de fundação da cidade justa, a Caverna constitui uma profunda crítica de Platão ao modo de vida dos cidadãos nas cidades gregas, nomeadamente na Atenas do seu tempo.

Num texto em que a descrição das cenas imaginadas é complementada por comentários esparsos, que colocam o diálogo na perspectiva psicopolítica, o filósofo denuncia a existência inautêntica dos seus concidadãos, defendendo a necessidade de usar a educação para corrigir as suas naturezas (514a1-2).

Descreve o comportamento de uma multidão aprisionada numa caverna, agrilhoadas de pernas e pescoço. Os prisioneiros são obrigados a contemplar sombras projetadas na parede da caverna por um fogo que arde nas suas costas e não podem ver, pelo fato de cada um ser forçado a manter a cabeça imóvel. As sombras reproduzem cenas representadas por homens que se movem ao longo de um passadiço, transportando artefatos diversos e ocasionalmente falando entre si. Deste modo, por não verem os próprios objetos, nem se verem uns aos outros, os prisioneiros julgariam que era sobre as sombras que conversavam e era delas que saíam as vozes.

### **A caverna**

A Caverna é proposta por analogia com qualquer comunidade grega que ocupa um território limitado, agrupando os cidadãos que nela vivem. Que vida é a deles? A alegoria mostra que consiste em assistirem ao

espetáculo ilusório proporcionado pela perpétua contemplação de sombras projetadas numa parede.

A força desta visão da vida na cidade imediatamente suscita uma reação do ouvinte/leitor. Será ela muito diferente da nossa? A um primeiro contato, o quadro descrito parece absurdo (em 517a8-c5, é fornecida a chave que permite descodificar a alegoria). No entanto, a pergunta que com ele Platão implicitamente faz a cada um dos seus leitores continua até hoje a ter sentido. Está cada um de nós seguro de que a sua é uma existência autêntica? Acaso se acha absolutamente certo da realidade do mundo em que o seu quotidiano o insere?

É na nossa própria incapacidade de responder a estas interrogações com uma afirmação clara e definitiva que reside o fascínio que a alegoria continua a exercer. E, no entanto, a continuação do texto vai mostrar que não é só a denúncia da nossa incapacidade de avaliação do quadro político em que estamos inseridos que a alegoria visa, mas ao que acontece como consequência dela.

Pensando que aquilo que é tomado pela realidade se reduz à sombra dos objetos percebidos, somos levados a perguntar que aconteceria, se — “soltos das cadeias e curados da sua ignorância” — os prisioneiros recuperassem a sua natureza (515c)? Deslumbrados por serem forçados a encarar subitamente o fogo, que diriam, se, ao serem confrontados com os próprios objetos, se dessem conta da ilusão em que tinham vivido? Acaso não poderiam supor que o que agora lhes era mostrado era menos real do que aquilo que antes viam (515d)?

### **Saída da caverna**

Contudo, se alguém forçasse um deles a sair da Caverna e a suportar a luz exterior, este teria de se habituar gradualmente, começando por olhar sombras e reflexos na água, para poder depois encarar as coisas (que ora temos por verdadeiras), e mais adiante o céu e os astros, de noite, até, por fim, vir a ser capaz de “contemplar o Sol” (516a-b). A partir desse

momento, essa contemplação haveria de o habilitar a compreender o lugar que o Sol ocupa no mundo visível, na posição de “responsável por todas as coisas” antes vistas (516c). Só então, tendo aí chegado, se sentiria capaz de tudo suportar, para não ter de voltar à sua existência anterior (516e).

### **Retorno à caverna**

Mesmo assim, talvez então perguntasse o que aconteceria se, por muito que rejeitasse a antiga vida na ilusão, quisesse regressar à Caverna. E aí perceberia que tudo havia de se passar tal como se tinha passado quando dela quis sair, pois, uma vez mais, deslumbrado<sup>16</sup>, mas agora, por ter subitamente abandonado a visão da luz do Sol, teria de suportar os mesmos vexames por parte do seus antigos companheiros, capazes até de ir ao extremo de o matarem, se tentasse soltá-los (516d-517a).

Chegado a esta conclusão, após uma breve síntese do processo de ascensão da alma ao inteligível, Sócrates detém-se numa derradeira consideração sobre a Ideia do bem. “Uma vez avistada”, no termo do inteligível, dela virá a compreensão da função que exerce no visível, como causa “de tudo o que é justo e belo”, criadora e senhora da luz, e, no inteligível, causa “da inteligência e da verdade” (517b-c).

Se aproximando do seu termo, a alegoria começa a se encaminhar para a problemática aludida de passagem no início do Livro VII (514a2), na qual o argumento do diálogo irá se concentrar até 535a. Entendida como “uma faculdade da alma”, a educação é aí mostrada como dependente do “órgão pelo qual [a alma] aprende”: “um olho” ... “que deve ser desviado, juntamente com a alma toda, das coisas que se alteram, até ser capaz de suportar a contemplação do ser” e “da sua parte mais brilhante”: o bem (518c).

Será apenas pela faculdade de pensar que esse órgão poderá ser orientado para onde deve (518d), se liberando do mundo mutável “para se

---

<sup>16</sup> As perturbações visuais sentidas nos momentos em que ocorrem súbitas variações de luminosidade prenunciam a concepção da educação como “translação” (*periatógê*) do visível para o inteligível (517e-518d, 533d).

voltar para a verdade” (519b). Como, porém, os que tiverem chegado aí renunciaram a tudo que a cidade lhes podia oferecer, como que transladados “para as ilhas dos Bem-aventurados” (519b-c), será necessário fazê-los descer novamente para junto dos prisioneiros, a fim de tirar partido deles “para a união da cidade” (519e-520a; mediante a prática da filosofia, se presume):

Deve, portanto, cada um por sua vez descer à habitação comum dos outros e habituar-se a observar as trevas. Com efeito, uma vez habituados, sereis mil vezes melhores do que os que lá estão e reconhecereis cada imagem, o que ela é e o que representa, devido a terdes contemplado a verdade relativa ao belo, ao justo e ao bom. E assim teremos uma cidade para nós e para vós, que é uma realidade e não um sonho, como actualmente sucede na maioria delas, onde combatem por sombras uns com os outros e disputam o poder, como se ele fosse um grande bem (520b).

### Síntese das três analogias

Embora se sobreponham umas às outras, Sol, Linha e Caverna não repetem um mesmo percurso<sup>17</sup>. As duas primeiras partem da experiência do visível, que separadamente submetem a análise. Por esse motivo, ao insistirem no recurso a metáforas visuais, começam por subsumir na visão a informação canalizada pelos outros sentidos, eventualmente processada por outras operações cognitivas. Como assinalámos atrás, a inflexão do “visível” para o “sensível” começa se manifestando em 510a, com a introdução da ‘opinião’, associada ao par “verdade/falsidade”. Dela resultará o recurso às “imagens”, nos passos em que antes se falava de “sombras”.

Na Caverna, esta caracterização será aprofundada, pois, enquanto os homens que se movem no passadiço carregam “imagens” e “formas” de artefatos, os prisioneiros acorrentados percebem-nas pelas “sombras” que

---

<sup>17</sup> Muitos intérpretes têm ensaiado a justaposição das três alegorias (ver V. Karasmanis, 1988, p. 147, nota 2-149; ca. Annas, 1981, p. 254 ss.). O problema não tem recentemente recebido muita atenção, mas ver Fine (1999, p. 226-46), que nos três símiles favorece uma análise de conteúdos, contra uma de objetos (p. 218-9, p. 231).

projetam, sendo sobre elas que emitem “opiniões”. No entanto, enquanto a Linha assimilava quatro tipos de operações psíquicas a outros tantos segmentos do real<sup>18</sup>, na Caverna é contemplada qualquer operação mental captada pelos sentidos.

Nos textos sobre o Sol e a Linha (até 510a), é pontual a referência qualificada a opiniões (510a10). No entanto, o texto permite que se pergunte como, no inteligível, poderão as “hipóteses” do ‘entendimento’ e do ‘saber’ ser avançadas, senão como “opiniões”? Opiniões, verdadeiras ou falsas, que, no *Mênon*, Sócrates sustenta se manifestarem na alma, mediante um interrogatório (85c), que cremos seja lícito associar ao exercício da reminiscência.

É a partir da ascensão — das sombras à Ideia do bem — que, na Caverna, é lançado o projeto platônico da educação. Como princípio cardinal da concepção é proposta a tese — fortemente crítica das práticas educativas atribuídas aos sofistas — de que não é possível “introduzir saber numa alma em que ele não existe”, tal como não se pode se introduzir “vista em olhos cegos” (518b-c).

Do princípio, decorre que a educação deve ser entendida como um processo de orientação das capacidades que a alma sempre teve para o objetivo que as três analogias concorreram em apontar como o mais elevado: “contemplar suficientemente o bem” (518c-519d). Todavia, a implementação do projeto depende de uma série de pressupostos teóricos, além de condições de aplicação.

Em primeiro lugar, o saber tem de estar na alma (*Mênon* 85e-86b; *Fédon* 72e-73a; 76e-77a). Depois, além de qualidades inatas e adquiridas, a alma tem de ser dotada do dinamismo necessário à implementação de um longo processo (*Carta VII* 340c-341a), tal como o currículo educativo terá de ser estruturado de modo a estimular a elevação da alma até ela ser capaz de refletir sobre “o que é” (*República VII* 521d3-525a).

---

<sup>18</sup> O texto só menciona a alma a partir do segmento inferior do inteligível, portanto, esta pode não se envolver nos contextos em que predomina a sensopercepção. Mas também será possível entender os dois segmentos do visível como meramente exemplificativos da relação entre “originais” e “sombras”.

Só no termo deste longo caminho — condensado num breve resumo da ascensão do prisioneiro na Caverna (530b-c) —, a alma se acha em condições de se dedicar à dialética (532e ss.). É para esse fim que, após um novo resumo, agora das disciplinas propedêuticas (533b-c), é desvelado o “método da dialética”. Avançando pela “destruição das hipóteses” (533d; ver VI 511b-c), ou seja, *descendo*, após a contemplação do bem, pela fixação das hipóteses em “verdades”, expressas em enunciados, o dialético se mostra como o único capaz de dizer o que é cada coisa (533b).

Será então possível fundir as três analogias:

Chamemos saber à primeira divisão, entendimento à segunda, crença à terceira e imaginário à quarta; opinião às duas últimas, inteligência às duas primeiras, sendo a opinião relativa à mutabilidade, e a inteligência à essência. E, assim como a essência está para a mutabilidade, está a inteligência para a opinião, e como a inteligência está para a opinião, está o saber para a crença e o entendimento para o imaginário<sup>19</sup> (533e-534a; ver VI 511d-e).

Após ter retomado os quatro segmentos nomeados na Linha, relacionando substratos ontológicos (essência e mutabilidade: *ousian*, contraposta a *genesin* — 534a3-4) com as duas seções epistemológicas da Linha (“inteligência” e “opinião”), subdivididas nas quatro operações psíquicas relativas a cada segmento (“saber”, “entendimento”, “crença” e “imaginário”), esta síntese esboça uma cadeia de proporções articuladas umas com as outras:

$$\begin{array}{ccccccccc} \text{essência} & & \text{inteligência} & & \text{inteligência} & & \text{saber} & & \text{entendimento} \\ \hline & = & & = & & = & & = & \\ \text{mutabilidade} & & \text{opinião} & & \text{opinião} & & \text{crença} & & \text{imaginário} \end{array}$$

Esta articulação, proporcionada pela sobreposição da Caverna à Linha e ao Sol, sintetiza a ontoepistemologia platônica<sup>20</sup>, exposta nos

<sup>19</sup> Inserimos, como já no Livro VI, pequenas modificações na tradução; “saber”, “crença”, “imaginário”, em vez de “ciência”, “fé”, “suposição”.

<sup>20</sup> Todavia, são muitas as dificuldades de obtenção de um encaixe perfeito entre as três analogias, sobretudo no que diz respeito aos chamados “objetos matemáticos” (White, 1979, p. 184-6; Annas, 1981, p. 272-4; Denyer, 2007, p.

Livros V-VII, da *República*. A prova de que o próprio Platão a reconhece como concepção autêntica pode se colher através da autocrítica que lhe é dirigida, no *Sofista* (247 ss.). Responsabilizando por ela os “amigos das Ideias”, o filósofo renuncia à cisão entre a comunhão do corpo com a “sensopercepção” na “mutabilidade” (*genesei*), e a da alma pensante com “o que o ser realmente é” (*pros tèn ontôs ousian*: 248a-b). Pouco depois, criticará indiretamente o monismo imobilista, que ele próprio atribui a Parmênides, quando enfaticamente o Hóspede de Eleia pergunta se: “... movimento, vida, alma e inteligência realmente não estão presentes no que de todo o modo é, que não vive nem pensa, mas venerável e sagrado, sem ter senso, permanece fixo e imóvel?” (248e-249a).

### Para lá da caverna

Para além de apontar a metodologia que permite o acesso à Ideia do bem, o ponto em que a Caverna decisivamente insiste é a de que só dois tipos de seres são reais (516a5-b2): no inteligível as Ideias, se elevando até à contemplação do bem; derivativamente, no sensível, os seres que constituem o mundo físico (516b4-7). Quanto aos outros, como mais não são do que imagens destes, as operações psíquicas a que se deve a sua captação servem só para permitir o acesso a eles, confirmando o princípio de que nenhuma captação será correta, enquanto consistir em imagens.

Todavia, os intérpretes dividem-se no que respeita ao êxito da metodologia proposta. Não há dúvida sobre o seu mérito, como procedimento que visa a elevar a alma ao inteligível. Mas nada garante que o exercício da dialética tenha permitido atingir o “princípio não-hipotético” condensado na Ideia do bem, de modo a operar a transformação das hipóteses em “verdades”. Sabemos que ‘dialético’ será aquele que é capaz de, de cada coisa, dizer “o que [ela] é” (533b2-3) e também “o que é capaz de uma “visão de conjunto” (*synoptikos*: 537c7;

---

302-8). A tese de que os “objetos matemáticos” existem como intermediários entre Ideias e sensíveis remonta a Aristóteles (*Met.* 16.987b14-18; 112.997b2; VII2.1028b19-21) e é confirmada por Proclo (*In Tím.* I, p. 61, Diehl).

537d enumera as qualidades nele patentes) do real. Mas não será certo que — mediante o método das perguntas e respostas — alguém tenha alguma vez chegado ao bem.

Outra alternativa será conferir à alma o poder de captar diretamente o saber, chegando ao bem por contato direto. Esse será o caminho de quantos defendem a “doutrina de Platão por uma tradição indireta” (M. Perine, 2009, 9-21; ver atrás, n. 11), ou aquele empreendido por Plotino e pelos muitos que, inspirados pelo Neoplatonismo, ultrapassam a lição platônica sobre o ser, para — além da ontologia — se projetarem na via conducente à henologia. Cremos ser possível encontrar nos diálogos indícios que consentem contemplar tal interpretação. Mas não encontramos lá provas de que Platão a tenha conseguido concretizar.

### **Epílogo: unidade e evolução da epistemologia platônica**

Lido da perspectiva da Filosofia Moderna, Platão é seguramente um racionalista, mas também um idealista e um realista. É racionalista e idealista porque constitui o conhecimento na alma e atribui ao pensamento a capacidade de o recuperar. É realista porque defende a realidade (existência) do sistema das Ideias, condensado na Ideia do bem: no ser ou para lá dele. Teremos, contudo, de conceder que há alguma contradição, ou pelo menos incompatibilidade, entre duas das leituras que a Modernidade atribui à filosofia platônica.

A interpretação aqui proposta tenta escapar a esta tensão ideológica abordando a obra do Mestre da Academia a partir do contexto cultural grego, nos sécs. V-IV, a. C. Considerando as três analogias analisadas atrás como o coração da epistemologia platônica, perguntamos que problemas elas abordam e como tentam resolvê-los. Encontramos pelo menos dois, sobre os quais a Obra do filósofo prolongadamente se debruça: os da unidade e infalibilidade do conhecimento.

O primeiro tem relevantes consequências no nosso modo de ler os diálogos, pois, tanto quanto podemos entender, falta neles aquilo que, a

partir de Descartes, entre outros aspectos relevantes, começou a se tornar evidente na literatura filosófica europeia: uma concepção unificada do conhecimento. Há várias explicações para essa falta, na filosofia clássica anterior a Aristóteles. À partida podemos justificá-la pelo fato de, nos textos que conhecemos na Língua Grega — tanto na arcaica como na clássica —, encontrarmos uma grande riqueza de expressões verbais, e em menor grau, a partir do séc. V, nominais, para descrever uma diversidade de operações cognitivas<sup>21</sup>.

Nas Poesias épica e didática, na Tragédia e na Comédia, são frequentemente atribuídas a deuses e divindades, e até a animais, capacidades cognitivas que só podem se compreender a partir daquilo que é a cognição para um ser humano. Por isso, será talvez lícito supor que esta amplitude não possa conviver com uma concepção unitária do conhecimento. Pois, se ‘conhecer’ é o mesmo para um homem e para um deus, será perfeitamente natural considerar infalível essa capacidade.

Talvez, por estes motivos, para se entregar sistematicamente ao estudo da cognição, Platão recorra a uma variedade de expressões cujos sentidos resistem a tentativas de definição e nas quais é difícil discernir relações internas. Note-se a concentração em passos canônicos (como *R. V 476-478*) de famílias de termos como: *gignôskô/gnôsis/gnôston/agnôia, epistamai/epistêmê, eidenai, ideô, oraô, blepô, akouô, nomizô, hageômai/hêgeomai, dianoeô/dianoia, dokeô/doxazô/doxa/doxaston, oida, manthanô*.

Que poderá o leitor extrair de traduções como — “conheço/conhecimento/cognoscível/ignorância, sei/saber, saber, sei, vejo, vejo/olho, ouço, acredito, creio, reflito, reflexão (entendimento), creio/opino/opinião/opinável, sei, compreendo” —, concentradas em duas páginas de texto grego<sup>22</sup>? Cremos que, para afrontar esta dificuldade, teremos de nos interrogar sobre os objetivos que Platão quer atingir, a

---

<sup>21</sup> Ver, no Canto II da *Ilíada*, a paradigmática expressão com que Homero se dirige às Musas, no início do Catálogo das Naus, pedindo auxílio para recordar os nomes e efetivos da armada dos Aqueus: “tudo sabeis porque presenciastes, enquanto nós só ouvimos a fama e nada sabemos”.

<sup>22</sup> Ainda por cima confundidas pelas muitas traduções recebidas de publicações noutras Línguas europeias.

partir do argumento do Livro V (Santos, 2020), com as analogias dos Livros VI-VII.

Respondendo ao primeiro quesito acima, criar condições que permitam definir a unidade do conhecimento mediante a defesa da possibilidade de trânsito entre as expressões com conteúdo cognitivo que os textos registram<sup>23</sup>. Quanto ao segundo quesito, não tanto para lhe responder, quanto para, de algum modo, mostrar a impossibilidade de o contornar, afrontar o obstáculo posto pela admissão de um único ‘conhecimento’, roubado aos deuses e passado aos mortais (*Protágoras* 320d-322a). Pois, a infalibilidade terá de lhe ser atribuída, uma vez que, como é implícito em Parménides, não será possível falar de um “não-conhecimento”, ou assumir a contradição, ao admitir que há um “conhecimento falso”.

Consequentemente, se entendermos o problema deste ponto de vista, a dificuldade não estará apenas na definição do conhecimento e na teoria que explica que e como é possível atingi-lo, mas também na integração nessa teoria da explicação do problema de como o erro será possível. E, com esta resposta, talvez nos seja permitido entrever a motivação que terá levado Platão a complementar a epistemologia comum ao *Fédon* e à *República* com o desenvolvimento do programa da epistemologia, na unidade dramática que liga o *Teeteto* ao *Sofista*.

## Referências

### A. Fontes:

ARISTÓTELES. *Metafísica de Aristóteles*. Edición trilingue por Valentin Garcia Yebra. Madrid: Gredos,1970.

PLATÃO. *Platonis Opera*. Ed. I. Burnet. IV. Oxford: Oxford University Press, 1900-1907.

---

<sup>23</sup> Particularmente pela articulação das famílias das raízes *epist-* e *gnô-* e  $\sqrt{\text{eid-}}$ ,  $\sqrt{\text{vid-}}$  com a variedade terminológica proporcionada pela ampliação de usos proporcionada pelas raízes *dok-*, *pha-*.

PLATÃO. **A República**. Tradução de M. H. da Rocha Pereira. Lisboa: Fundação Calouste Gulbenkian, 2001<sup>9</sup>.

PLATÃO. **O Sofista**. Tradução de H. Murachco, J. Maia Jr., J. T. Santos. Lisboa: Fundação Calouste Gulbenkian, 2011.

PLATO. **Republic**, with an English translation by Paul Shorey, I-II, Heinemann. London: Harvard University Press, 1935.

PLATO. **Republic**. Trad. english Paul Shorey, I-II, Heinemann. London: Harvard University Press, 1935.

PLATON. **La République**. Texte établi et traduit par Émile Chambry. Les Belles Lettres. Paris 1933.

PLATONE. **La Repubblica V**. Traduzione e commento a cura di Mario Vegetti. Bibliopolis, Napoli 2003.

## **B. Livros e artigos:**

ALLEN, R. E. (ed.). **Studies in Plato's metaphysics**. London: Routledge & Kegan Paul, 1965.

ALLEN, R. E. Participation and Predication in Plato's Middle Dialogues. **The Philosophical Review**, 69, No. 2, p. 147-164, 1960 [Studies in Plato's Metaphysics, 1965, p. 43-60].

ANNAS, J. **An introduction to Plato's Republic**. Oxford: Oxford University Press, 1981.

CROSS, A.C. & WOOLEY, A. D. **Plato's Republic: A Philosophical Commentary**. MacMillan, 1964.

DENYER, N. Sun and Line: the Role of the Good. In: FERRARI, G. R. F. **Plato's Companion to the Republic**. Cambridge: Cambridge University Press, 2007, p. 284-309.

EL MURR, D. Why the Good? Appearance, Reality and the Desire for the Good, in *Republic* VI 504D-506B. **Méthexis**, XXVII, p. 47-60, 2014.

- FERRARI, G. (ed.). **Plato's companion to the Republic**. Cambridge: Cambridge University Press, 2007.
- FINE, G.(ed.). **Plato 1: metaphysics and epistemology**. Oxford: Oxford University Press, 1999a.
- FINE, G. Knowledge and Belief in *Republic* 5–7. In. G. Fine (ed.). **Plato 1**. Oxford: Oxford University Press, 1999b, p. 215-46.
- FINE, G. (ed.). **The Oxford handbook to Plato**. Oxford: Oxford University Press, 2011.
- GUTHRIE, W. **History of greek philosophy IV**. Cambridge: Cambridge University Press, 1975.
- HARTE, V. Plato's Epistemology. In: **The Oxford handbook to Plato**, 1911, p. 191-216.
- KARASMANIS, V. Plato's Republic: The Line and the Cave. **Apeiron**, 21, p. 147-71, 1988.
- KRAUT, R. (ed.). **The Cambridge companion to Plato**. Cambridge: Cambridge University Press, 1992.
- OWEN, G. E. L. A Proof in the *Peri ideôn* **Journal of Hellenic Studies**, 77, p. 103-11, 1957.
- PERINE, M. Leituras entre o escrito e o não-escrito. In: PERINE, M. (org.) **Estudos platônicos: sobre o ser e o aparecer, o belo e o bem**. São Paulo: Loyola, 2009, p. 9-21.
- RAVEN, J. E. **Plato's thought in the making**. Cambridge: Cambridge University Press, 1965.
- REALE, G. **Per una nuova interpretazione di Platone**. 20. ed. Bompiani, Milano, 1997 [Para uma nova interpretação de Platão. São Paulo: Loyola, 2004].
- SANTAS, G. X. The Form of the Good in Plato's *Republic*. **Philosophical Inquiry**, 2, 1, p. 374-403, 1980 [abbreviated version In: Fine (ed.). **Plato 1**, 1980, p. 247–74].
- SANTAS, G. X. Knowledge and Belief in Plato's *Republic*. **Boston Studies in the Philosophy of Science**, 121, Kluwer Academic Publishers, p. 45-59, 1990.

Santos, J. T. O argumento sobre “as coisas que são e não são”: R. V 476A1-479B10.

**Prometheus**, 32, p. 1-12, 2020

SZLEZÁK, T. A. **Platon und die Schriftlichkeit der Philosophie**. De Gruyter, Berlin, 1985

[**Platão e a escritura da filosofia: Análise de estrutura dos diálogos da juventude e da maturidade à luz de um novo paradigma**. São Paulo: Loyola, 2009].

VEGETTI, M. L'idea del bene nella Repubblica di Platone. *Discipline Filosofiche* 1, p. 207-

29, 1993.

VLASTOS, G. Reasons and Causes in the *Phaedo*. In: VLASTOS, G. **Platonic Studies**<sup>2</sup>.

Princeton: Princeton University Press, 1981, p. 43-57.

VLASTOS, G. Degrees of Reality in Plato. In: VLASTOS, G. **Platonic Studies**<sup>2</sup>. Princeton:

Princeton University Press, 1981a, p. 58-75.

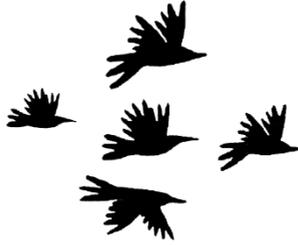
WHITE, N. P. **Plato on Knowledge and Reality**. Indianapolis: Hackett, 1976.

WHITE, N. P. **A companion to Plato's Republic**. Oxford: Basil Blackwell, 1979.

WHITE, N. P. Plato's Metaphysical Epistemology. In: KRAUT, Richard. **The Cambridge**

**Companion to Plato**. Cambridge: Cambridge University Press, 1992, p. 277-310.

WIELAND, W. **Platon und die Formen des Wissens**. Göttingen 1982, p. 159-85.



## Como livrar-se da lógica e ser feliz

*Desidério Murcho*<sup>1</sup>

Considere-se o seguinte argumento:

Marguerite é romancista.

Logo, Yourcenar é romancista.

O argumento não é válido. Contudo, suponha-se que Marguerite é Yourcenar. Neste caso, parece não haver maneira de a premissa ser verdadeira quando a conclusão é falsa, dado que, afinal, trata-se da mesma pessoa com dois nomes diferentes. Este género de exemplos foi usado por Edgington (2004) e parece sugerir que um argumento é dedutivamente válido se e só se:

1. Não é possível que a sua conclusão seja falsa se todas as premissas forem verdadeiras.
2. Consegue-se saber *a priori* que 1.

---

<sup>1</sup> Estudou na Universidade de Lisboa e no King's College London e atualmente é Professor de Filosofia na Universidade Federal de Ouro Preto (UFOP). Publicou, dentre outros, os livros *A Existência de Deus: O Essencial*, *Lógica: O Essencial*, *Lógica Elementar: Raciocínio, Linguagem e Realidade*, *Todos os Sonhos do Mundo e Outros Ensaios*, *Filosofia em Directo*, *Pensar Outra Vez: Filosofia, Valor e Verdade*, *O Lugar da Lógica na Filosofia e Essencialismo Naturalizado: Aspectos da Metafísica da Modalidade*, entre outros. Organizou ainda os volumes *A Ética da Crença*, *Viver Para Quê?* Ensaios Sobre o Sentido da Vida e *Enciclopédia de Termos Lógico-Filosóficos* (coorganizador), entre outros. E-mail: [desiderio@ufop.edu.br](mailto:desiderio@ufop.edu.br). Crítica – Revista de Filosofia e Ensino (ISSN: 1749-8457): <https://criticanarede.com/>

1 é a condição metafísica, 2 a epistémica. De notar que isto se alarga facilmente às verdades lógicas. Uma afirmação é uma verdade lógica se e só se:

1. É necessariamente verdadeira.
2. Consegue-se saber *a priori* que 1.

Este é o sentido alargado de “verdade lógica”, segundo o qual a frase “Nenhum particular completamente azul é verde” recebe essa designação. Para captar o sentido estrito, formal, basta acrescentar outra condição:

3. Obedece-se à condição 2 usando apenas métodos formais de prova.

Assim, a frase “Yourcenar é romancista se for romancista” é uma verdade lógica no sentido estrito: é necessariamente verdadeira e consegue-se saber *a priori* que o é usando apenas métodos formais de prova. O mesmo acontece no caso do sentido estrito, formal, de validade dedutiva; o argumento seguinte é válido neste sentido:

Marguerite é romancista.  
 Marguerite é Yourcenar.  
 Logo, Yourcenar é romancista.

Este é um primeiro passo promissor na direção de uma explicação da validade dedutiva e da verdade lógica na lógica clássica e noutras lógicas próximas; mas não se aplica no caso de verdades lógicas ou quase lógicas como as dos indexicais: “Estou aqui” é talvez conhecida *a priori* e é algo como uma verdade lógica, mas é defensável que é uma verdade contingente e não necessária. Deixemos de lado estes casos, para já.

Tirando a lógica dos indexicais, a proposta inspirada em Edgington (2004) é compatível com as lógicas clássicas e a maior parte das não-clássicas. Além disso, é de algum modo pré-teórica, no sentido de não depender de pressupor que algum género de teoria dos modelos ou de teoria da demonstração é a maneira de definir ou compreender a validade e a

verdade lógica. Assim, é o tipo de coisa que Field (2015) favorece: é uma tentativa, aceitável para lógicos que defendem diferentes perspectivas, de definir a validade. Field pensa que um tratamento da validade em termos de teoria dos modelos ou de teoria da demonstração “não pode possivelmente ser o que queremos dizer com ‘válido’” (Field, 2015, p. 33), sobretudo porque dado qualquer tratamento — o clássico, por exemplo —, o lógico rival concordará que uma dada inferência é válida naquele tratamento particular, ainda que pense que não é genuinamente válida: “mesmo quem rejeita a lógica frásica clássica, irá concordar que as inferências frásicas que o lógico clássico aceita são válidas *neste* sentido” (Field, 2015, p. 33), ou seja, no sentido de ser classicamente válida. Duvido que este género de argumento seja bem-sucedido quanto ao que Field pretende, porque um lógico relativista partidário da teoria dos modelos ou da teoria da demonstração irá insistir que não há validade fora de qualquer sistema dado de teoria lógica. Contudo, apesar de a perspectiva aqui apresentada não pretender montar para já qualquer género de argumentação contra o relativista partidário da teoria dos modelos ou da demonstração, pretende-se que tenha o género de neutralidade teórica que Field apoia:

Uma maneira de tentar explicar o conceito de validade é defini-lo noutros termos (mais familiares ou mais básicos). Como vimos, qualquer tentativa de usar a teoria dos modelos ou a teoria da demonstração para este fim seria vã; mas há uma alternativa proeminente de tentar defini-la. Na sua forma mais simples, a validade é explicada dizendo que uma inferência (ou argumento) é válida se e só se preserva a verdade por necessidade lógica (Field, 2015, p. 35).

A preservação necessária da verdade num argumento válido é talvez o mesmo que a impossibilidade de ter conclusão falsa se todas as premissas forem verdadeiras; nesse caso, a proposta de Field é muito próxima da mencionada leitura de Eddington (2004). Ele não parece dar-se conta, contudo, que, sem uma cláusula epistémica, precisa ou de excluir o necessário *a posteriori*, ou de declarar que o argumento “Marguerite é romancista; logo, Yourcenar é romancista” é, afinal, válido.

Começemos pela primeira alternativa: suponha-se que não se aceita a ideia de verdades necessárias *a posteriori*. Nesse caso, não se acredita que há argumentos inválidos que obedecem à condição 1, acima. Até certo ponto, esta é uma discordância quanto à extensão dos conceitos de verdade necessária e de verdade lógica, e não acerca dos próprios conceitos. Quem defende este ponto de vista acredita que aqueles conceitos são coextensionais (excluindo talvez a lógica dos indexicais e outras lógicas não clássicas relacionadas), e isso equivale à rejeição das verdades necessárias *a posteriori*.

Poderá não parecer, mas, caso se distinga cuidadosamente as componentes epistémicas das metafísicas no que diz respeito à verdade lógica, esta tese da coextensionalidade não é fácil de defender. Não parece certamente razoável que nos limitemos a pressupô-la. E por que razão haveria o necessariamente verdadeiro de estar tão intimamente em harmonia com o que conseguimos saber *a priori*? A menos, é claro, que a tese seja que “verdade lógica” *significa* apenas “conhecível *a priori*, de maneira puramente lógica”. Nesse caso, contudo, a tese não é apenas que os conceitos de verdade necessária e conhecimento *a priori* são coextensionais; a tese é que é o mesmíssimo conceito, sob diferentes apresentações. Esta sugestão será discutida daqui a pouco. O ponto importante aqui é que não é muito promissor rejeitar o necessário *a posteriori*, ao mesmo tempo que se aceita que os conceitos de verdade necessária e de conhecimento *a priori* são diferentes; caso se queira rejeitar o necessário *a posteriori*, há outra maneira mais promissora de o fazer, como se verá.

Para já, suponha-se que os conceitos de verdade necessária e de conhecimento *a priori* são muito diferentes, ainda que sejam coextensionais. Nesse caso, está disponível uma redução eliminativa dos conceitos de verdade lógica e de validade dedutiva – na formulação de Field (2015), uma definição “noutros termos (mais conhecidos ou mais básicos)”. Uma verdade lógica não é senão uma verdade necessária que conseguimos conhecer *a priori*; um argumento dedutivamente válido não

é senão uma impossibilidade — é impossível que a conclusão seja falsa e as premissas verdadeiras — que somos capazes de conhecer *a priori*. Dado um mundo com árvores e átomos e tudo isso, e agentes cognitivos capazes de conhecimento *a priori*, já se tem verdades lógicas e argumentos dedutivamente válidos. Talvez exista também validade e verdade lógica num domínio de deuses com linguagens formais e objetos abstratos, mas há uma maneira alternativa, pedestre, de compreender esses conceitos. A relação nuclear de implicação, a preservação necessária da verdade, ou seja o que for que se queira usar para falar da validade dedutiva, não é, assim, algo para lá do velho mundo empírico de árvores e rios e das suas relações, mais um certo gênero de funções cognitivas de agentes como nós; e o mesmo se aplica às verdades lógicas. Por outras palavras, a modalidade lógica é redutível à metafísica modal mais o *a priori*. Não há modalidade lógica independente.

Isto levanta uma interrogação, contudo, acerca da natureza das próprias verdades necessárias. Porém, por que razão haveria isto de ser assim tão diferente das verdades comuns, que não são necessárias? Suponha-se que a necessidade diz realmente respeito ao mundo, e não ao nosso conhecimento, crenças ou outras funções cognitivas. Nesse caso, a afirmação “A água é necessariamente  $H_2O$ ”<sup>2</sup> é verdadeira se for isso que a água é necessariamente, tal como a afirmação “Russell escreveu *Os Problemas da Filosofia*” é verdadeira se ele escreveu aquele livro. Se o conceito de necessidade for tomado como um conceito metafísico genuíno e não como um conceito epistémico disfarçado, nada há de especial com respeito a, pelo menos, algumas verdades necessárias; algumas têm

---

<sup>2</sup> Note-se que a tese do químico de que a água é  $H_2O$  não é a tese de que aquilo a que se chama “água” em contextos informais é apenas  $H_2O$ ; ao invés, é a tese de que há muito de  $H_2O$  no que se chama “água” nesses contextos, e talvez também que é esta composição química que explica muitas das suas características químicas. Claro que a água que se bebe tem muitas outras coisas, muitas delas importantes para os interesses humanos, e é claro também que os tomates e os bebês são maioritariamente feitos de água, mas essa não é a característica mais relevante que têm (sobre isto, veja-se Abbott, 1999). Em qualquer caso, a tese do metafísico é apenas que a tese do químico, se for verdadeira, é-o necessariamente, e isto é compatível com a ideia de que os químicos erraram completamente e a água não é realmente  $H_2O$ . Seria algo surpreendente, contudo, que aparecesse um argumento filosófico ou linguístico bem-sucedido contra um resultado bem investigado da química.

veridadores, tal e qual como algumas verdades não-necessárias<sup>3</sup>. Quando um químico examina uma porção de água, talvez não veja que é necessariamente  $H_2O$ . Contudo, também não vê que é  $H_2O$ , se com “ver” se quer dizer algo de direto e óbvio. Para *concluir* que aquela porção de água é  $H_2O$ , o químico percorre um conjunto complexo de testes e de inferências. Mas suponha-se que o químico poderia limitar-se a olhar e ver que aquela porção de água é  $H_2O$ , mas que não conseguia ver que o é necessariamente. Como conseguiria isto ser uma razão para acreditar que a água é contingentemente  $H_2O$ ? A menos que se comece por afirmar que a necessidade é, de algum modo, um conceito epistémico, está longe de ser óbvio que todas as verdades necessárias se devam ver facilmente; por que razão haveria isso de ser assim?

Por isso, pressupondo que a necessidade é um conceito metafísico genuíno, que diz respeito à realidade em geral e não apenas às maneiras humanas de a compreender, há uma maneira promissora de explicar a verdade lógica e a validade dedutiva. O mundo faz existir verdades necessárias, e torna impossível que algumas afirmações sejam falsas se outras forem verdadeiras; e, depois, os seres humanos conseguem conhecer *a priori* algumas dessas necessidades e impossibilidades. Talvez estas duas teses precisem de ser clarificadas<sup>4</sup>. Mas nenhuma é terrivelmente problemática. Talvez a primeira seja apenas uma característica primitiva do mundo, algo que se explica mas não em termos de outros aspectos ainda mais primitivos do mundo. Acrescentando o operador de necessidade, consegue-se gerar a afirmação “A água é necessariamente  $H_2O$ ”, a partir da afirmação mais simples “A água é  $H_2O$ ”; mas isso não é uma boa razão para acreditar que o mundo começa por ter

---

<sup>3</sup> Suponha-se que alguém acredita que só o presente é real e que as verdades necessárias são sempre verdadeiras. Nesse caso, essa pessoa acredita que algumas verdades necessárias não *têm* veridadores — ainda que venham a tê-los no futuro. Contudo, esta é uma questão geral sobre qualquer verdade acerca do futuro, e não especialmente acerca de verdades necessárias: ainda que todas as verdades sejam contingentes ou até *simpliciter* (no sentido não serem metafisicamente contingentes, nem necessárias), esta dificuldade com os veridadores — se é que existe — permanece.

<sup>4</sup> Nomeadamente, para responder à preocupação de que talvez se precise de factos negativos. Contudo, note-se, primeiro, que esta é uma dificuldade também com verdades não-modais e, segundo, que a cláusula metafísica de Eddington (2004) pode ser parafraseada sem a preocupante palavra “impossível”: necessariamente, se as premissas de um argumento válido são verdadeiras, também a sua conclusão o é.

água que é simplesmente  $H_2O$ , e depois outra característica — ou mundo possível — torna isso necessário. Em geral, não é de esperar que o mundo seja assim tão solícito, que espelhe todos os aspectos da linguagem — ao invés, são os seres humanos que tentam representar características do mundo nas suas linguagens, e nem sempre o fazem de maneiras que não sejam enganadoras. Quanto à segunda tese, seja o que for que se pense do conhecimento *a priori*, é preciso acomodar o género de justificação usada na matemática e é preciso aceitar que isso é muito diferente do género de justificação usada na química. E isto é suficiente para prosseguir. Ainda que se rejeite o *a priori* como termo filosófico de arte, e ainda que se pense que não há aqui qualquer distinção epistémica clara e útil, esta perspectiva é aceitável, porque certamente se aceita que os métodos comuns de demonstração da matemática são diferentes, ainda que não radicalmente, dos métodos da química.

Livrámo-nos da lógica, se a entendermos deste modo? De certo modo, sim. Não há relações lógicas independentes, propriamente ditas; há apenas o mundo com as suas características e agentes cognitivos capazes de entender algumas delas de uma maneira particular. As verdades lógicas são-no exclusivamente porque são necessidades metafísicas que se consegue conhecer *a priori*. Não se trata de as necessidades metafísicas o serem por serem verdades lógicas. A modalidade metafísica e a epistémica são primitivas relativamente à modalidade lógica; as primeiras duas explicam a última, e não ao contrário. Repare-se que isto não significa que se fica obrigado a rejeitar a modalidade lógica como conceito primitivo. Talvez seja preciso esse conceito para levar a cabo outras tarefas, ou talvez existam outras razões para não nos livrarmos da lógica como conceito primitivo. Tudo o que o raciocínio até agora faz é expandir um pouco uma maneira promissora de explicar a verdade lógica usando apenas conceitos não-lógicos.

E se alguém acreditar que para se ter um argumento válido ou uma verdade lógica a condição 1 acima é perfeitamente suficiente? Esta não é uma perspectiva particularmente promissora, porque “Marguerite é

Yourcenar” seria, então, uma verdade lógica; e é óbvio que isto parece errado. Caso se aceite o necessário *a posteriori*, esta afirmação é vista como necessariamente verdadeira, se for verdadeira; mas por que razão seria também logicamente verdadeira? E quais são as regras de inferência de uma teoria lógica segundo a qual qualquer frase da forma  $a = b$  seja uma verdade lógica? E como se distinguiria aqueles casos em que uma afirmação aparentemente da mesmíssima forma lógica é falsa, como “Quine é Kripke”? Em suma, a crença de que a condição 1 é suficiente para explicar a verdade lógica e a validade dedutiva é implausível a menos que se rejeite o necessário *a posteriori*. É de supor que a motivação dessa perspectiva é a ideia de que a verdade lógica e a verdade necessária andam de mãos dadas; mas a melhor maneira de ir por essa via é rejeitar o necessário *a posteriori*, rejeitando assim que “Marguerite é Yourcenar” é necessariamente verdadeira se for verdadeira, e rejeitando também que o argumento que parte de “Marguerite é romancista” e conclui que “Yourcenar é romancista” seja válido. Esta sugestão será desenvolvida daqui a pouco, mas façamos primeiro uma breve pausa.

O que se viu até agora é que um entendimento redutor da validade e da verdade lógica não é pelo menos completamente inusitado, desde que se leve a sério o conceito metafísico de verdade necessária, juntamente com um conceito qualquer de conhecimento *a priori* suficientemente rico para dar conta do conhecimento matemático. O que se verá de seguida é uma alternativa a esta perspectiva, apropriada para quem foge a sete pés das necessidade metafísicas. E um dos pontos deste artigo é sugerir a ideia de que estas são as únicas duas perspectivas promissoras que há — as perspectivas rivais acabam por cair numa ou na outra destas duas.

Regressemos a indexicais como “Estou aqui”. Suponha-se que alguém acredita que isto é uma espécie de verdade lógica. A condição 1 acima não é, aparentemente, respeitada: não é necessário que eu esteja aqui se eu poderia estar noutro lado. E quando a condição 1 não é respeitada, a 2 também não o é (acontecendo o mesmo à condição 3). Contudo, suponha-se que alguém acredita que os indexicais deste género

são verdades lógicas. A característica epistêmica óbvia de afirmações indexicais como aquela é que são *a priori*: seja por meios exclusivamente linguísticos, seja exclusivamente pelo raciocínio, sou capaz de saber que estou aqui. De modo que talvez uma perspectiva mais liberal do conceito de verdade lógica exija apenas que sejamos capazes de saber *a priori* que uma afirmação é verdadeira. Talvez na maior parte dos casos a necessidade esteja também envolvida, mas não em todos.

Isto sugere uma maneira redutora diferente, e muito minimalista, de entender a verdade lógica e a validade dedutiva. Começemos pela primeira:

**Uma afirmação é uma verdade lógica se e só conseguimos saber a *priori* que é verdadeira.**

Suponha-se por instantes que isto é levado a sério. Nesse caso, há uma maneira de devolver algo a quem deplora o necessário *a posteriori*: o conceito de necessidade metafísica já não faz falta.

Recorde-se que uma maneira de fazer oposição ao necessário *a posteriori* se exprime de tal modo que a disputa diz respeito exclusivamente à extensão dos conceitos. Isto não é plausível, entre outras razões, porque não oferece qualquer razão plausível, que não seja *ad hoc*, para rejeitar o necessário *a posteriori*. Contudo, esta dificuldade desaparece caso se elimine o próprio conceito de necessidade metafísica. Não é uma questão de afirmar que seja o que for que só se consegue saber *a posteriori* é contingente por razões muitíssimo misteriosas, mas antes que os filósofos têm usado desde sempre os conceitos vazios de necessidade e contingência: verdade seja dita, nada no mundo responde a tais conceitos. Não se trata de a água ser contingentemente H<sub>2</sub>O porque não se consegue saber *a priori* que é necessariamente H<sub>2</sub>O, mas antes que a água é apenas H<sub>2</sub>O, e acabou-se a conversa. Não há modalidades metafísicas, de todo em todo; só há modalidades epistêmicas.

Uma razão poderosa para acreditar que há verdades necessárias *a posteriori* emerge do seguinte resultado lógico simples. Tome-se dois nomes que denotem o mesmo particular, como “Marguerite” e “Yourcenar”, e forme-se a frase “Marguerite é Yourcenar”. Acrescente-se, como segunda premissa, a verdade lógica de que Marguerite é Marguerite. Pressuponha-se agora uma forma irrestrita da lei de Leibniz (a parte da indiscernibilidade de idênticos), que permita a substituição de “Marguerite” por “Yourcenar” em qualquer contexto não-epistêmico. A conclusão surpreendente é que Marguerite é necessariamente Yourcenar. Esta conclusão é inabalável, pressupondo que o conceito de necessidade aqui em questão seja metafísico, e não epistêmico. Em qualquer contexto não-epistêmico, a lei de Leibniz permite realmente a substituição irrestrita de idênticos por idênticos. Se o conceito não for metafísico, contudo, há boas razões para bloquear a inferência. Afinal, há boas razões para bloquear a inferência que chega a “Galeno acredita que Yourcenar é romancista” partindo das premissas “Marguerite é Yourcenar” e “Galeno acredita que Marguerite é romancista”. De modo que, caso se comece por pressupor que a necessidade não é senão um conceito epistêmico, não há razão para aceitar a substituição irrestrita de idênticos por idênticos, sem a qual não há qualquer prova do necessário *a posteriori*. De “Marguerite é Yourcenar” e de “Marguerite é necessariamente Marguerite” não há uma rota válida que conclua “Marguerite é necessariamente Yourcenar”, caso “necessariamente” não seja lido metafisicamente (ou de maneira transparente, se se quiser), porque nesse caso o uso irrestrito da lei de Leibniz é inválido.

O curioso nisto é que mesmo antes de se começar a explorar a ideia de que a necessidade não é senão um conceito epistêmico disfarçado — o *a priori*, na verdade — há uma razão independente para suspeitar que o conceito de verdade lógica não é puramente metafísico. Pois se o fosse, as seguintes inferências seriam ambas indisputavelmente válidas, mas a segunda certamente que não o é:

É verdadeiro que Marguerite é romancista.

Marguerite é Yourcenar.

Logo, é verdadeiro que Yourcenar é romancista.

É logicamente verdadeiro que Marguerite é Marguerite.

Marguerite é Yourcenar.

Logo, é logicamente verdadeiro que Marguerite é Yourcenar.

Seja porque se considera que “logicamente verdadeiro” é uma espécie de operador metalinguístico que exige aspas, seja por outra razão, o resultado é que se a restrição ao uso da lei de Leibniz é uma indicação de que o conceito envolvido não é puramente metafísico, então parece que a sugestão de Eddington (2004) está na pista certa: o conceito de verdade lógica, ao contrário do de verdade *simpliciter*, talvez tenha conteúdo epistémico. O eliminativista modal só dá mais um (grande) passo e propõe-se eliminar completamente os conceitos de verdade lógica e de necessidade metafísica, reduzindo-os exclusivamente ao *a priori*. Deste ponto de vista, os filósofos foram todos vítimas de uma maneira de falar, tomando a verdade lógica e a verdade necessária como conceitos de pleno direito, quando afinal não passam de maneiras diferentes de falar do que se consegue saber *a priori*. Repare-se que agora não há qualquer mistério quanto à harmonia delicada entre o necessário e o *a priori*, porque um dos conceitos acabou de ser deitado pela janela fora. Não é que existam dois conceitos não-vazios – a verdade lógica e a necessidade metafísica – que acontece serem misteriosamente coextensionais; trata-se antes de haver só um conceito – o *a priori* – sob duas vestes diferentes.

Quão promissor é isto? É certamente terrivelmente minimalista. Toma dois conceitos importantes envolvidos num volume considerável de trabalhos filosóficos e formais e propõe-se deitá-los fora, relendo-os como um conceito epistémico sob diferentes aparências. Será que funciona?

Uma preocupação inicial é que o conceito de verdade necessária parece crucial para explicar a validade dedutiva; pois tome-se os seguintes argumentos:

Rute e Dorothy são filósofas.

Logo, Rute é filósofa.

Rute ou Dorothy são filósofas.

Logo, Rute é filósofa.

Suponha-se agora que elas são as duas filósofas. Sob esta condição, os dois argumentos têm premissas verdadeiras e conclusões também verdadeiras. Contudo, o primeiro é válido, coisa que o segundo não é. A diferença, é claro, é que, no primeiro caso, não é possível que a conclusão seja falsa, caso a premissa seja verdadeira, mas isto é exatamente o que é possível no segundo caso. De modo que o conceito modal de possibilidade desempenha aqui um papel proeminente. Como explicaria o eliminativista esta diferença?

Bem, a diferença é que no primeiro caso, mas não no segundo, há maneiras de saber *a priori* que se as premissas são verdadeiras, também a conclusão o é. Assim, deste ponto de vista, a relação entre as premissas e a conclusão de um argumento válido explica-se como se segue:

É dedutivamente válido concluir *B* de *A* se e só se há maneiras de saber *a priori* que se *A* então *B*<sup>5</sup>.

Deste ponto de vista, a diferença entre os dois argumentos acima é epistémica, e não metafísica. E isso parece funcionar.

O eliminativista propõe assim uma dupla redução. A necessidade metafísica é completamente eliminada, no sentido de não existir tal coisa; só há verdades lógicas e validades dedutivas. Os conceitos de verdade lógica e de validade dedutiva são, então, reduzidos ao *a priori* mais o velho conceito de verdade *simpliciter*, ou seja, verdade sem quaisquer modalidades metafísicas. Segundo o eliminativista modal, não há

---

<sup>5</sup> Repare-se que se evita aqui deliberadamente termos técnicos e semitécnicos, como “implicação”. Além disso, a condicional material clássica é o que se visa em “se *A* então *B*”; talvez a proposta funcione também com maneiras não-clássicas de entender a condicional, mas se isso não acontecer, insira-se “não-*A* ou *B*” no lugar apropriado.

diferença metafísica entre as verdades lógicas e as outras; há uma diferença epistémica importante, e é tudo.

Porém, não se esconde aqui uma modalidade metafísica qualquer? Que se quer dizer ao afirmar que há maneiras de saber algo *a priori*? Suponha-se que vim a saber *a posteriori* que o último teorema de Fermat é verdadeiro, ao ler sobre isso e ao assistir até a uma palestra sobre o caso. Contudo, não sei provar realmente este resultado. De modo que o meu conhecimento parece *a posteriori*. Contudo, há maneiras de saber *a priori* que o último teorema de Fermat é verdadeiro — reconstruindo a demonstração depois de a estudar, por exemplo. Imagine-se, contudo, que nunca fiz tal coisa. Nesse caso, dizer que há maneiras de saber *a priori* que o último teorema de Fermat é verdadeiro mais não faz, ao que parece, do que esconder a contrafactual modal de que eu poderia tê-lo sabido *a priori*.

Quão séria é esta dificuldade? Claro que se se revelar que há aqui algum género de discurso modal escondido e não-eliminável, a eliminação proposta do conceito de necessidade metafísica fracassa. Contudo, o eliminativista tem uma resposta imediata. Dizer que eu poderia ter sabido algo *a priori* é apenas, habitualmente, uma maneira de dizer que há outros que o sabem *a priori* — de modo que há maneiras de nós o sabermos *a priori*. De facto, quando se olha para o conhecimento supostamente *a priori* que uma pessoa tem de algo mais exigente do que uma verdade lógica ou matemática simples, vê-se que depende imenso do conhecimento *a posteriori* adquirido em contacto com o trabalho dos outros. De modo que, na maior parte dos casos, dizer que sei algo *a posteriori* que poderia ter sabido *a priori* talvez queira realmente dizer que outra pessoa o sabe *a priori*. Numa situação radical, em que um agente cognitivo como nós está sozinho e sem acesso ao trabalho alheio, está longe de ser óbvio que poderia ter sabido *a priori* algo que de facto sabe *a posteriori*. Nesta situação radical, parece que não há coisa alguma que o agente poderia ter sabido *a priori* mas que sabe realmente *a posteriori*.

O ponto mais importante, contudo, é que a exigência que se faz ao eliminativista modal talvez não seja razoável. Isto porque caso se leve a

sério a ideia de que os conceitos modais são afinal epistémicos, são epistémicos até ao fim. Em geral, o eliminativista traduz as caixas e diamantes das lógicas modais aléticas como se segue:

□A: há maneiras de saber *a priori* que A.

◇A: não há maneiras de saber *a priori* que não-A.

Assim, dizer que eu poderia ter sabido *a priori* que *p* traduz-se como se segue: é possível que eu saiba *a priori* que *p*, ou seja, não há maneiras de eu saber *a priori* que não sei *a priori* que *p*. Sim, isto faz tropeçar a língua, mas isso não é uma boa razão para rejeitá-lo.

De modo que talvez o eliminativismo modal não enfrente qualquer desafio imediato na frente epistémica. E que dizer da frente metafísica? Do ponto de vista eliminativista, as verdades lógicas são apenas verdades comuns que se consegue conhecer *a priori*. As verdades comuns, contudo, pelo menos algumas delas, parece que têm veridadores diretos: a afirmação “Russell nasceu na Inglaterra” é verdadeira porque Russell nasceu na Inglaterra, e é este acontecimento que é responsável pela sua verdade. Dado que o eliminativista modal afirma que metafisicamente não há qualquer diferença entre as verdades lógicas e as não-lógicas — a diferença é meramente epistémica — parece que pelo menos algumas verdades lógicas têm veridadores, e, além disso, são precisamente como os veridadores das verdades comuns, não-lógicas. Como pode ser tal coisa?

A resposta é que o veridador de uma verdade lógica como “Ou Russell nasceu na Inglaterra, ou não” é o mencionado facto de ele ter nascido na Inglaterra. Isto porque dada qualquer afirmação da forma *p* ou *q*, o que for responsável pela verdade de uma das afirmações (seja *p*, seja *q*), é também responsável pela verdade da afirmação maior<sup>6</sup>. Não é preciso um domínio especial de factos, características formais ou entidades abstratas; basta o mundo empírico comum, e depois agentes linguísticos como nós que

---

<sup>6</sup> Este género de perspectiva é defendida por Heathcote (2003).

criam linguagens com certas características. Estas características linguísticas não são responsáveis, só por si, pela verdade dessas afirmações; o responsável é o mundo. Mas são talvez essas características que permitem saber *a priori*, sem olhar para o mundo, que são verdadeiras. Uma vez mais, a distinção cuidadosa entre o que diz respeito ao mundo em geral e o que diz respeito àquela parte do mundo que são os agentes, com as suas linguagens, crenças e outras funções cognitivas, desempenha aqui um papel de relevo.

De notar que a perspetiva eliminativista é muito diferente em substância, ainda que não superficialmente *in verba*, da ideia de que as verdades lógicas são de algum modo meras convenções linguísticas. As convenções linguísticas têm realmente um papel a desempenhar nas verdades lógicas; mas também, no caso das verdades que não são lógicas, isso acontece. Afinal, sob convenções linguísticas diferentes, a afirmação “Russell nasceu na Inglaterra” é falsa, em vez de verdadeira — uma convenção, por exemplo, segundo a qual “nasceu” quer dizer “foi vítima de homicídio”. Com respeito à linguagem, a convenção tem *sempre* um papel a desempenhar. Contudo, o eliminativista modal não afirma que com respeito às verdades lógicas, as convenções linguísticas, *só por si*, são responsáveis pela sua verdade; o mundo é responsável pela sua verdade, e as convenções linguísticas são, na melhor das hipóteses, responsáveis pela nossa capacidade para conhecê-las *a priori*. O único sentido em que as convenções linguísticas são responsáveis pelas verdades lógicas é o mesmíssimo sentido em que são responsáveis pelas verdades não-lógicas: sem convenções linguísticas não haveria afirmações, e se estas forem os portadores primitivos de valor de verdade, não haveria verdades, de todo em todo. Isto não significa, contudo, que as convenções linguísticas são veridadores; na melhor das hipóteses, são responsáveis pela existência de portadores de verdade.

Suponha-se agora que alguém defende que o mundo não é responsável pela verdade de “Ou Russell nasceu na Inglaterra, ou não” porque esta afirmação é verdadeira independentemente do que acontece

no mundo. É difícil, contudo, dar sentido a este desafio. Dadas as convenções linguísticas que temos e o mundo tal como é, a afirmação “Russell nasceu na Inglaterra” é verdadeira. O mesmíssimo verificador desta afirmação é também responsável pela verdade de quaisquer afirmações como “Russell nasceu na Inglaterra ou as baleias não têm barbatanas” e “Russell nasceu na Inglaterra ou todos os carros são azuis”. Em geral, dado um verificador responsável pela verdade de uma afirmação  $p$ , esse verificador é também responsável pela verdade de  $p$  ou  $q$ . Uma diferença importante no caso de afirmações da forma  $p$  ou  $\text{não-}p$  é que sabe-se *a priori* que são verdadeiras. Considere-se uma afirmação como “Há extraterrestres”. Não se sabe se esta afirmação é verdadeira. Sabe-se *a priori*, contudo, que é verdadeira se houver extraterrestres e falsa caso contrário. No mesmo espírito, sabe-se *a priori* que “Ou há extraterrestres ou não” é verdadeira. Para que esta afirmação seja verdadeira, contudo, é suficiente que existam extraterrestres, e é também suficiente que não existam. Ou seja, aconteça o que acontecer, isso torna aquela afirmação verdadeira. Por que razão seria isto uma dificuldade? Admita-se que toda a gente que está numa sala é louro. A afirmação “Alguém naquela sala é louro” é verdadeira, independentemente da pessoa que se tenha em mente; qual é a dificuldade aqui? É preciso que esteja lá alguém que não é louro para que os que são louros tornem aquela afirmação verdadeira? Isto não parece promissor. Sim, a afirmação está, digamos, sobredeterminada: é verdadeira independentemente da pessoa que se tiver em mente. Mas isto não parece uma dificuldade séria.

De modo que o eliminativismo modal não parece enfrentar dificuldades óbvias. E oferece a quem quiser rejeitar o necessário *a posteriori* — quem acredita que se uma afirmação não é uma verdade lógica, então não é uma verdade necessária — uma maneira promissora de o fazer de maneira apropriada; pois suponha-se que alguém acredita que as verdades necessárias e as verdades lógicas são coextensionais e que essa pessoa não se livrou do próprio conceito de verdade necessária. Nesse caso, essa pessoa não parece ter uma maneira que não seja circular de

rejeitar o uso irrestrito da lei de Leibniz que prova que existem verdades necessárias que não são verdades lógicas. Por outro lado, essa pessoa terá de admitir que tudo o que é logicamente contingente é realmente contingente — de modo que o sucesso da ciência parece tornar-se um milagre. Sempre que um cientista prevê um resultado, qualquer coisa pode acontecer. Talvez ele deixe cair uma bola para provar uma tese sobre a gravidade, e a bola transforma-se num elefante cor-de-rosa. Por que razão não haveria isso de acontecer? Tudo o que não é uma contradição lógica é possível.

Suponha-se que se reage a isto afirmando que “possível” quer apenas dizer “logicamente possível”, e que isso não é uma possibilidade *real*. Bem, acredita essa pessoa ainda na coextensionalidade entre a verdade lógica e a verdade necessária, e que se trata de dois conceitos diferentes? Nesse caso, note-se que isto implica que tudo o que não é uma verdade lógica é contingente, e isto não quer dizer que é apenas logicamente contingente; quer dizer que é *realmente* contingente. Em contraste, o eliminativista modal livra-se completamente da necessidade e da contingência metafísicas, de modo que ser logicamente contingente não quer dizer que é contingente em qualquer sentido metafisicamente relevante. É por isso que esta perspectiva parece a melhor aposta para quem suspeita das necessidades não-lógicas.

Outra preocupação, talvez mais promissora, é que não se consiga explicar o mundo sem algum tipo de modalidade metafísica — talvez porque não há qualquer maneira não-modal que seja promissora de compreender as leis da natureza, por exemplo. Talvez isso seja assim, apesar de ser de duvidar até alguém ter realmente tentado desenvolver, sem sucesso, com os instrumentos filosóficos e lógicos que hoje temos, a velha ideia humiana de que talvez o mundo seja apenas o raio de uma coisa depois da outra<sup>7</sup>. Apesar disso, suponha-se que a hipótese humiana não tem pernas para andar. Nesse caso, o eliminativismo modal é falso. Note-

---

<sup>7</sup> Não se afirma aqui que Hume o pensava; e é claro que ele pensava que a causalidade era crucial para compreender o mundo. Ele pensava, contudo, que era muito difícil justificar as nossas crenças causais.

se, contudo, que a dificuldade aqui é metafísica, e não conceptual. Conceptualmente, a redução dos conceitos lógicos ao *a priori* (mais a verdade *simpliciter*) não parece enfrentar dificuldades sérias; acontece apenas que se o mundo não for humiano, o eliminativismo modal é falso.

Parece, pois, que há duas maneiras promissoras e não equivalentes de explicar a verdade lógica e a validade dedutiva, seja reduzindo-as ao *a priori* mais a necessidade metafísica, seja reduzindo-as ao *a priori* mais a verdade *simpliciter*. Em ambos os casos, a verdade lógica e a validade dedutiva revelam-se conceitos que afinal não são primitivos; são construídos com base noutros mais primitivos – o género de coisa que Field (2015) procura. Isto parece promissor por duas razões. Primeiro, como se viu brevemente, nenhuma das abordagens parece enfrentar dificuldades óbvias iniciais. Segundo, as outras abordagens parece que fazem mais ou menos o mesmo – mas não de maneira tão óbvia. Por exemplo, imagine-se que alguém favorece a teoria da demonstração ou a teoria dos modelos como via para explicar a verdade lógica e a validade dedutiva. Parece que há duas maneiras de entender qualquer uma destas abordagens. Ou se acredita que os modelos modelam alguma coisa no mundo em geral, ou se pretende cortar todos os laços entre os modelos e o mundo em geral. E o mesmo acontece no caso da teoria da demonstração: os passos de uma derivação e as regras de inferência podem ser vistas exclusivamente como dispositivos formais independentes do mundo em geral, ou como algo que representa a validade dedutiva corriqueira.

Comecemos pela ideia de que os modelos e as demonstrações não são sobre o mundo em geral. Talvez sejam sobre um mundo abstrato, platónico? Nesse caso, é-se empurrado à mesma para a perspectiva de Eddington (2004), inserindo esse mundo abstrato no lugar do mundo em geral de rios, galáxias e átomos; e será preciso incluir, em qualquer caso, a capacidade humana para saber *a priori* o que se passa nesse mundo platónico. Além disso, parece que é preciso aceitar que este mundo platónico tem alguma relação com o nosso mundo pedestre, pois é

verdadeiro que Russell nasceu na Inglaterra ou não. De modo que, no fim das contas, é preciso regressar à Terra, não passando o mundo abstrato de uma paragem temporária. Qualquer perspectiva plausível que não exija essa paragem parece estar, quanto a este aspecto, em vantagem.

De modo que talvez as coisas sejam mais promissoras caso se corte as amarras entre as linguagens formais e o mundo em geral, rejeitando-se também que tais linguagens formais sejam acerca de um mundo platónico; não são acerca seja do que for, talvez? Neste caso, o *locus* da verdade lógica não é as afirmações corriqueiras como aquela acerca de Russell, mas antes afirmações formais como “ $p \vee \neg p$ ”, e estas não são acerca seja do que for — não são acerca de um mundo platónico, nem são acerca do mundo em geral. Isto não é particularmente promissor, porque parece que uma sequência de fórmulas como esta representa um número indefinido de afirmações corriqueiras que têm aquela forma lógica. “Russell nasceu na Inglaterra” tem valor de verdade, mas “ $p$ ” não tem tal coisa: é apenas um lugar a ser preenchido por diferentes afirmações, algumas das quais serão verdadeiras, mas não todas. E se “ $p$ ” não tem valor de verdade, “ $p \vee \neg p$ ” também não é verdadeira, quanto mais logicamente verdadeira. Apesar de tudo, suponha-se que alguém insiste que pretende realmente cortar todas as amarras entre as linguagens formais e o mundo quotidiano. Nesse caso, como se haverá de compreender a explicação que essa pessoa oferece das verdades-corriqueiras-aparentemente-lógicas? Ainda que não sejam Verdades Lógicas Reais porque não se exprimem usando uma linguagem formal, é certo que exigem explicação. E essa pessoa não nos oferece explicação alguma. Contudo, assim que começar a fornecer uma explicação, o mundo em geral e as nossas faculdades *a priori* irão aparentemente voltar a ter um papel a desempenhar, e toda aquela conversa formal não passa, de novo, de uma paragem temporária.

De maneira que livrarmo-nos completamente das modalidades lógicas talvez seja a via a seguir; o *a priori* mais a verdade — seja necessária, seja *simpliciter* — parecem conseguir explicar, por si, a verdade

lógica e a validade dedutiva. E quanto à felicidade? Fica para outra ocasião. Este foi apenas o primeiro passo lógico<sup>8</sup>.

## Referências

ABBOTT, B. Water = H<sub>2</sub>O. **Mind**, 108.429, p. 145-8, 1999.

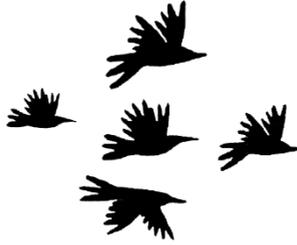
EDGINGTON, D. Two Kinds of Possibility. **Aristotelian Society Supplementary**, Volume 78, p. 1-22, 2004.

FIELD, H. What is Logical Validity? In: CARET, Colin R. & HJORTLAND, Ole T. **Foundations of Logical Consequence**. Oxford: Oxford University Press, 2015, p. 33-70.

HEATHCOTE, A. Truthmaking and the Alleged Need for Relevance. **Logique et Analyse** 46.183/184, p. 345-64, 2003.

---

<sup>8</sup> Agradeço a Matheus Silva e a Pedro Merluzzi os generosos comentários construtivos que me permitiram melhorar este trabalho. A versão inglesa original deste artigo foi publicada no volume *Logic, Intelligence, and Artifices: Tributes to Tarcisio H. C. Pequeno*, ed. Jean-Yves Beziau, Francicleber Ferreira, Ana Teresa Martins e Marcelo Pequeno (Londres: College Publications, 2018), com o título “How to Get Rid of Logic and Be Happy”.



## Princípio da explosão e outras esquisitices lógicas

*Frank Thomas Sautter*<sup>1</sup>

*Vamos nos acostumar a viver na estranheza, na esquisitice,  
protegendo-nos como podemos de atos, fatos e ideias bizarros.*

Lya Luft

### Introdução

Na Lógica, como na vida, estranhezas, esquisitices e bizarrices podem ser toleradas, desde que justificadas e controladas para que não afetem o curso normal da vida e dos pensamentos.

Reuni, aqui, seis esquisitices lógicas, cujo denominador comum é dado pelo binômio implicação e vacuidade. Utilizo “implicação” no sentido peirciano amplo de cópula lógica, ou seja, como um termo polissêmico interpretado, em alguns casos, como consequência lógica e, em outros casos, como operador de condicional (SÁNCHEZ VALENCIA, 2004, p. 474). A vacuidade ocorre quando um valor lógico distinguido<sup>2</sup> é obtido em condições

---

<sup>1</sup> Doutor em Filosofia pela Universidade Estadual de Campinas e Professor Titular do Departamento de Filosofia da Universidade Federal de Santa Maria.

<sup>2</sup> Utilizo uma adaptação da terminologia das lógicas polivalentes, em que os valores são divididos em distinguidos – aqueles que são objeto de nosso interesse – e não-distinguidos – aqueles que não são objeto de nosso interesse. Essa adaptação se deve ao fato de que, como será mostrado imediatamente a seguir, não apenas proposições estão sujeitas ao fenômeno da vacuidade, mas também argumentos e, eventualmente, outras entidades lógicas.

atípicas<sup>3</sup> em decorrência da ausência de algo. Por exemplo, uma proposição categórica universal afirmativa é vacuamente verdadeira (o valor distinguido) quando não há instâncias do termo sujeito (a ausência de algo), independente do termo predicado (condição atípica)<sup>4</sup>; um argumento é vacuamente válido (o valor distinguido) quando não há circunstâncias em que as premissas são simultaneamente verdadeiras (a ausência de algo), independente da conclusão (condição atípica). Nesses exemplos, a semântica da “implicação” é um fator decisivo para o surgimento da vacuidade: a semântica da condicional material, no primeiro exemplo, a semântica da consequência lógica clássica, no último caso.

Na primeira seção, apresento uma esquisitice da Lógica Tradicional, em que a discussão é centrada na existência ou inexistência de instâncias de termos gerais. Na segunda seção, trato de um exemplo recente – o Princípio do Bêbado – em que a discussão gira em torno de domínios do discurso vazios ou não vazios. Na terceira seção, examino uma discussão importante do século XIX em torno do Paradoxo da Barbearia, a partir do ponto de vista de Lewis Carroll. Na quarta seção, investigo uma esquisitice lógica vinculada a duplas “implicações” (bicondicionais materiais), retirada da obra literária *Ardil-22*. Na quinta seção, discuto um teorema existencial leibniziano em Teoria Moral, segundo o qual há aquele que ama todas as pessoas boas, e que depende também da exigência da (não-) vacuidade do domínio do discurso. Finalmente, na sexta seção, examino uma das principais esquisitices da Lógica Contemporânea, a saber, o Princípio da Explosão.

## 1. Do pressuposto existencial à vacuidade na lógica tradicional

As proposições categóricas da Lógica Tradicional são formalizadas na Lógica Contemporânea por fórmulas da forma  $\forall x (\varphi x \supset \psi x)$  (universal afirmativa),  $\forall x (\varphi x \supset \neg \psi x)$  (universal negativa),  $\neg \forall x (\varphi x \supset \neg \psi x)$

---

<sup>3</sup> Naturalmente o que é típico e o que é atípico variam de caso a caso.

<sup>4</sup> O mesmo ocorre com uma proposição categórica universal negativa: ela é vacuamente verdadeira quando não há instâncias do termo sujeito, independente do termo predicado.

(particular afirmativa) e  $\neg\forall x (\varphi x \supset \psi x)$  (particular negativa). Essa formalização tem o efeito de retirar a validade universal das conversões por limitação, assim como das oposições contrária, subcontrária e subalterna, que são vitais à Teoria da Prova aristotélica<sup>5</sup>. A solução para reestabelecer o que foi retirado é “pressupor” existência, o que, em termos da Lógica Contemporânea, expressa-se mediante a verdade de  $\exists x Px$ , para todo predicado P<sup>6,7</sup>. Por exemplo, se não houvesse pressuposição existencial, proposições categóricas em oposição contrária – as universais afirmativa e negativa de mesmo termo sujeito e mesmo termo predicado – poderiam ser simultaneamente verdadeiras por vacuidade, ou seja, poderiam ser verdadeiras porque o termo sujeito não se aplica a nada.

Poderia ser objetado que o problema (ou a solução, conforme o ponto de vista) não está propriamente na pressuposição existencial, mas no comportamento lógico das condicionais tácitas nas proposições categóricas universais. Para responder a isso, é instrutivo examinar o procedimento para a determinação do valor-de-verdade de proposições categóricas universais. Considere-se o caso da proposição categórica universal afirmativa “Todos os filósofos são taciturnos”. Ela expressa a subordinação do conceito “ser um filósofo” ao conceito “ser taciturno”. Aplica-se a ela o seguinte procedimento:

- a) Isola-se, no universo do discurso, a coleção dos filósofos.
- b) Isola-se, no universo do discurso, a coleção dos taciturnos.

---

<sup>5</sup> Por exemplo, a conversão por limitação (*conversio per accidens*), indicada na mnemotécnica medieval pela letra “P”, é utilizada na redução (perfectibilização) de quatro modos válidos, a saber, DARAPTI, FELAPTON, BRAMANTIP e FESAPO (ver GARRIDO, 1983, p. 160). Já a oposição contrária é, elegantemente, utilizada na prova de modos inválidos (ver SAUTTER, 2019, p. 146).

<sup>6</sup> Na Lógica Tradicional, o pressuposto existencial do termo geral T pode ser imposto pela verdade de “Algum T é T” ou, equivalentemente, pela falsidade de “Nenhum T é T”.

<sup>7</sup> A pressuposição de existência é legítima para todos os predicados (termos gerais) formados por abstração e para alguns predicados (termos gerais) formados por combinação lógica de predicados (termos gerais) previamente disponíveis. Desse modo, a pressuposição de existência é legítima para todos os predicados (termos gerais) “simples” e também para alguns, mas não todos, predicados “complexos”.

Em termos extensionais, se a diferença da coleção dos filósofos em relação à coleção dos taciturnos for vazia, a proposição é verdadeira; caso contrário, ela é falsa.

Mas, o que dizer do complemento, em relação ao domínio do discurso, da coleção dos filósofos? Ele é simplesmente irrelevante nesse procedimento! A forma como a relação lógica fundamental de subordinação é expressa na Lógica Contemporânea<sup>8</sup>, mediante quantificador universal e condicional, impõe que a semântica da condicional seja tal que ela é verdadeira quando seu antecedente é falso. A condicional material é a resposta correta para que casos irrelevantes não afetem o teste do valor-de-verdade de fórmulas que expressam a subordinação de conceitos.

A estranheza causada pela semântica da condicional material se deve ao fato dela ser examinada no plano proposicional, abstraída da quantificação. Mas, se se quiser manter constante a interpretação da condicional na passagem da Lógica Proposicional para a Lógica Quantificacional<sup>9</sup>, a semântica da condicional material é a correta.

A despeito de seus méritos, a solução oferecida pela condicional material tem o efeito colateral de fazer certas fórmulas vacuamente verdadeiras, ou seja, verdadeiras porque o antecedente de uma condicional (material) é falso, o que causa estranheza ao público não-especializado.

## 2. Princípio do bêbado

Uma esquisitice lógica contemporânea é o Princípio do Bêbado. Smullyan (1978, p. 209) o caracteriza do seguinte modo: há ao menos uma

---

<sup>8</sup> A Teoria do Silogismo Categórico é uma teoria da subordinação de conceitos: as universais expressam a subordinação de conceitos e as particulares expressam a não-subordinação de conceitos.

<sup>9</sup> Nas suas obras sobre os fundamentos da Teoria dos Números, Frege criticou aqueles que concebiam, por exemplo, a operação de soma de números reais como uma extensão da operação de soma de números naturais. O erro dessa abordagem reside no fato de que números naturais são utilizados para contar, enquanto que números reais são utilizados para medir; e se eles são objetos para distintas utilizações, no mínimo, deve ser admitida a possibilidade de que tenham fundamentos distintos.

pessoa tal que se ela bebe, então todos bebem. Que esse princípio é verdadeiro prova-se por casos do seguinte modo:

*Caso 1:* Todas as pessoas bebem. Seja  $k_1$  uma pessoa qualquer<sup>10</sup>. Nesse caso, “Se  $k_1$  bebe, então todas as pessoas, inclusive  $k_1$ , bebem.” é verdadeira porque antecedente e consequente são verdadeiros. Por generalização existencial, há ao menos uma pessoa tal que se ela bebe, então todos bebem.

*Caso 2:* Alguma pessoa não bebe. Seja  $k_2$  uma pessoa qualquer que não bebe<sup>11</sup>. Nesse caso, “Se  $k_2$  bebe, então todas as pessoas, inclusive  $k_2$ , bebem.” é *vacuamente* verdadeira porque o antecedente (e o consequente) é (são) falso(s). Novamente, por generalização existencial, há ao menos uma pessoa tal que, se ela bebe, então todos bebem.

Se o domínio do discurso é constituído por pessoas e o predicado unário “Bx” é interpretado como “x bebe”, o Princípio do Bêbado pode ser expresso pela seguinte fórmula:  $\exists x (Bx \supset \forall y By)$ . A validade universal dessa fórmula é contra-intuitiva, mas a Figura 1 apresenta uma demonstração por tablôs dessa validade universal em domínios não-vazios. Nas expansões das linhas 2 e 5, utilizamos parcialmente a regra da negação da condicional, e foram necessários dois parâmetros para o fechamento do tablô.

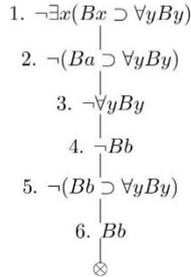
Contudo, em domínios vazios, essa fórmula não é válida. Hailperin (1953) e Quine (1954) fornecem a regra para as situações em que domínios vazios são admitidos<sup>12</sup> e, conforme a mesma, a invalidade da fórmula é imediata.

---

<sup>10</sup> Esse passo de prova depende, evidentemente, da não vacuidade do universo do discurso.

<sup>11</sup> Novamente, esse passo de prova depende da não vacuidade do domínio do discurso.

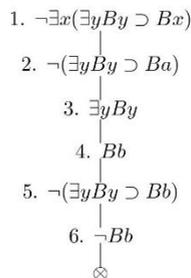
<sup>12</sup> Hailperin (1953, p. 197) expressa a regra e a justifica do seguinte modo: “[...] for empty domains (x)p means the conjunction of a zero number of conjunctands and hence has the value T [...].  $(\exists x)p$  is to have the value F in empty domains.” Quine (1954, p. 177) a expressa do seguinte modo: “An easy supplementary test enables us anyway, when we please, to decide whether a formula holds for the empty domain. We have only to mark the universal quantifications as true and the existential ones as false and apply truth-table considerations.”



**Figura 1.** Validade universal (em domínios não-vazios) de  $\exists x (Bx \supset \forall y By)$ .

A seguinte formulação do Princípio do Bêbado é válida mesmo em domínios vazios: Se, e somente se, houver ao menos uma pessoa, então há ao menos uma pessoa tal que se ela bebe, então todos bebem. Ela pode ser formalmente expressa pela seguinte fórmula:  $\exists x (x = x) \equiv \exists y (By \supset \forall z Bz)$ .

Smullyan (1978, p. 210) também fornece a seguinte versão “dual” do Princípio do Bêbado: há ao menos uma pessoa tal que se alguém bebe, então ela bebe. Utilizando as anteriores convenções, o “dual” do Princípio do Bêbado pode ser expresso pela seguinte fórmula:  $\exists x (\exists y By \supset Bx)$ . A validade universal dessa fórmula é contra-intuitiva, mas a Figura 2 apresenta uma demonstração por tablôs dessa validade universal em domínios não-vazios. Novamente, nas expansões das linhas 2 e 5, utilizamos parcialmente a regra da negação da condicional, e foram necessários dois parâmetros para o fechamento do tablô.



**Figura 2.** Validade universal (em domínios não-vazios) de  $\exists x (\exists y By \supset Bx)$ .

A despeito do que Smullyan diz, o que se entende por dualidade em lógica é tal que a negação é a sua própria dual, os quantificadores universal e existencial são duais um do outro e o dual da condicional é a negação da condicional com antecedente e consequente trocados. A dual de uma fórmula é falsa em todas as interpretações em que a fórmula é verdadeira, e somente nessas. Nesse sentido, a real dual da fórmula  $\exists x (Bx \supset \forall y By)$  é a fórmula  $\forall x (\exists y By \wedge \neg Bx)$ , e ela é universalmente inválida em domínios não-vazios, conforme demonstrado no tablô da Figura 3.

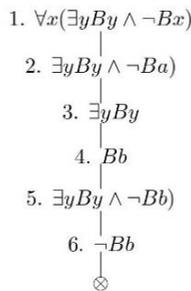


Figura 3. Invalidez universal (em domínios não-vazios) de  $\forall x (\exists y By \wedge \neg Bx)$ .

### 3. Paradoxo da barbearia

O Paradoxo da Barbearia (“On the way to the barber-shop”) está no centro de uma controvérsia a respeito da condicional, cujos atores principais são John Cook Wilson e Lewis Carroll, e que perdurou de 1892 a 1905, ou seja, sete anos após a morte de Lewis Carroll, John Cook Wilson ainda estava publicando material sobre o assunto (BARTLEY, III, 1986, p. 445).

Lewis Carroll discutiu ao menos seis versões do paradoxo (BARTLEY, III, 1986, p. 449-465). Lindemann e Sautter (2019, p. 100-104) discutem detalhadamente o paradoxo e a solução dada por Lewis Carroll. Em sua formulação mais simples, dada ao Paradoxo da Barbearia por Lewis Carroll, ele se apresenta na seguinte versão abstrata (BARTLEY, III, 1986, p. 451):

Há duas proposições, A e B.

Seja admitido que:

Se A é verdadeira, B é verdadeira. (i)

Seja uma outra Proposição C, tal que:

Se C é verdadeira, então se A é verdadeira, B não é verdadeira. (ii)

Nemo e Outis diferem sobre a verdade de C.

Nemo diz que C não pode ser verdadeira; Outis diz que ela pode ser.

Nemo representa o ponto de vista de Wilson e Outis, o ponto de vista de Carroll. O núcleo da discussão concentra-se na compatibilidade ou não de pares de proposições da forma *Se p, então q* e *Se p então não-q*. Outis dá à condicional a sua interpretação usual, ou seja, a semântica da condicional material, o que é rejeitado por Nemo.

Se interpretamos as condicionais do par como condicionais materiais, então elas são verdadeiras quando o antecedente é falso, ou seja, elas são compatíveis. A “vacuidade” aqui é dada pelo fato de se obter um valor lógico distinguido (o verdadeiro) quando há ausência de circunstâncias que fazem do antecedente uma proposição verdadeira (ausência de algo), independente do conseqüente (condição atípica).

#### 4. Catch-22

“Catch-22” ou “Ardil-22” é um romance de Joseph Heller (1961) no qual há uma regra militar, cujo nome fornece o título da obra, que indica as “condições” e a lógica por detrás da dispensa de missões aéreas perigosas. O seguinte trecho, em tradução própria, contém todos os elementos necessários para tratarmos de uma versão muito peculiar de vacuidade (as marcações entre colchetes são minhas e serão utilizadas na reconstrução do argumento):

Havia somente um ardil e este era o Ardil 22, que dizia ser a preocupação com a própria segurança, em face de perigos reais e imediatos, o processo de uma mente racional [b]. Orr estava louco e podia ser dispensado. Tudo o que ele tinha a fazer era pedir [c]; mas, assim que o fizesse, não estaria mais louco e teria de voar em novas missões. Orr seria louco se voasse em novas missões,

e são se não o fizesse, mas se ele estivesse são, ele teria de voar em novas missões [a] (HELLER, 1961, p. 41).

Goldstein (2004) reconstrói o trecho acima sob forma de bicondicionais. Se o domínio do discurso é constituído por militares, o predicado unário “Px” for interpretado como “x pode evitar missões aéreas perigosas”, o predicado unário “Ix” for interpretado como “x é insano”, e o predicado unário “Dx” for interpretado como “x solicita dispensa de participar de missões aéreas perigosas”, as condições de Catch-22 são as seguintes (GOLDSTEIN, 2004, p. 299-300):

- a)  $\forall x (Px \equiv Ix)$ , ou seja, um militar pode evitar missões aéreas perigosas se, e somente se, for insano;
- b)  $\forall x (Ix \equiv \neg Dx)$ , ou seja, um militar é insano se, e somente se, não solicitar dispensa de participar de missões aéreas perigosas;
- c)  $\forall x (Dx \equiv Px)$ , ou seja, um militar pode evitar missões aéreas perigosas se, e somente se, solicitar dispensa de missões aéreas perigosas.

Por contraposição e transitividade, é trivial demonstrar que:

- d)  $\forall x (Px \equiv \neg Px)$ , ou seja, um militar pode evitar missões aéreas perigosas se, e somente se, não pode evitar missões aéreas perigosas.

Goldstein (2004, p. 300) denomina a bicondicional de “vácua”, porque ela não expressa condição nenhuma. Ele a contrasta com bicondicionais que *meramente* não podem ser satisfeitas como a seguinte: “Você não pode evitar missões aéreas perigosas se, e somente se, você puder trisectar um ângulo arbitrário usando apenas régua e compasso.”<sup>13</sup> (GOLDSTEIN, 2004, p. 300)

---

<sup>13</sup> A triseção de um ângulo arbitrário apenas com régua (não marcada) e compasso é um problema clássico da Antiguidade, e sua solução é que é impossível fazê-lo.

## 5. Teorema de Leibniz

Leibniz define boa pessoa como aquela que ama todas as pessoas. Se o domínio do discurso é constituído por pessoas, o predicado unário “ $Bx$ ” for interpretado como “ $x$  é bom” e o predicado binário “ $Axy$ ” for interpretado como “ $x$  ama  $y$ ”, a definição de boa pessoa pode ser expressa pela seguinte fórmula:  $\forall x (Bx \equiv \forall y Axy)$ . Dessa definição pode-se concluir o “irrealista” Teorema de Leibniz: há ao menos uma pessoa que ama todas as pessoas boas<sup>14</sup>. Esse teorema prova-se, por casos, do seguinte modo:

*Caso 1:* Há, ao menos, uma pessoa boa. Seja  $b$  uma pessoa boa. Por definição,  $b$  ama a todos, inclusive todas as pessoas boas. Por generalização existencial, há uma pessoa que ama todas as pessoas boas.

*Caso 2:* Não há nenhuma pessoa boa. Sejam  $b$  e  $c$  pessoas (que não são boas pessoas)<sup>15</sup>. Segue-se que “Se  $b$  é uma pessoa boa, então  $c$  ama  $b$ ” é uma proposição verdadeira, porque o antecedente é falso, e é falso porque não há pessoas boas<sup>16</sup>. Por generalização universal, qualquer que seja a pessoa, se ela é uma boa pessoa,  $c$  a ama. Por generalização existencial, há uma pessoa que ama todas as pessoas boas.

Utilizando as convenções anteriores, esse teorema pode ser expresso pela seguinte fórmula:  $\exists y \forall x (Bx \supset Ayx)$ . O tablô da Figura 4 apresenta a validade do argumento cuja única premissa é a definição leibniziana de boa pessoa e cuja conclusão é o Teorema de Leibniz. Nas expansões das linhas 4 e 10 utilizamos parcialmente a regra da negação da condicional, e foram necessários três parâmetros para o fechamento do tablô.

---

<sup>14</sup> Evidentemente o candidato natural para aquele que ama todas as pessoas boas é a “superpessoa” Deus. E é também evidente que, de um ponto de vista teológico, essa tese está aquém de Deus: ele ama todas as pessoas *simpliciter*.

<sup>15</sup> Esse passo de prova depende, evidentemente, da não vacuidade do universo do discurso.  $b$  e  $c$  não precisam ser, necessariamente, pessoas distintas.

<sup>16</sup> Utilizei essa repetição para destacar que a situação se enquadra em minha caracterização de fenômeno de vacuidade.

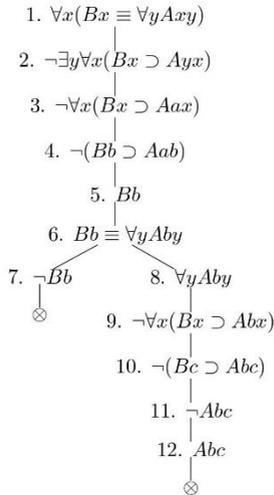


Figura 4. Validade do Argumento de Leibniz

## 6. Princípio da explosão

Finalmente, a suprema das esquisitices lógicas: o Princípio de Explosão.

Na sua formulação mais geral, em termos de consequência semântica, o Princípio de Explosão asserir que de um conjunto inconsistente de proposições segue-se qualquer proposição, ou seja, no contexto da Lógica Clássica, ele asserir que qualquer proposição pode ser sustentada com base em um conjunto de proposições que não podem ser simultaneamente verdadeiras. Em termos de argumentos, se a coleção das premissas é inconsistente, o argumento é vacuamente válido, qualquer que seja a sua conclusão.

Newton da Costa (1993) foi o pioneiro em apresentar “lógicas alternativas” que derrogavam o Princípio de Explosão; ele é um dos pais fundadores das Lógicas Paraconsistentes, cujo foco é lidar de maneira não-clássica com conjuntos inconsistentes de proposições. Aqui examinarei, brevemente, duas alternativas ao Princípio de Explosão, muito menos sofisticadas do que as propostas por Newton da Costa, mas, também, do meu ponto de vista, mais intuitivas e também mais próximas da Lógica

Clássica do que as soluções apresentadas por Newton da Costa e sua Escola.

Uma alternativa à aceitação de quaisquer consequências lógicas de um conjunto inconsistente de proposições é adotar um ponto de vista diametralmente oposto, a saber, a supressão total das consequências lógicas de um conjunto inconsistente de proposições, ou seja, nenhuma consequência lógica se segue de um conjunto inconsistente de proposições. Sautter (2008) desenvolveu uma “lógica proposicional” em que uma relação de consequência lógica que é clássica para os conjuntos consistentes e que é nula para os conjuntos inconsistentes de proposições.

A supressão total de consequências, no caso de inconsistência, tem o apoio de pensadores ilustres. George Berkeley, por exemplo, ofereceu um *rationale* para a supressão total na seguinte passagem de *O analista* (1734):

Nenhuma Conclusão lícita poder ser diretamente extraída de duas Suposições inconsistentes [entre si], nada há de mais manifesto do que isso. Você pode, de fato, supor qualquer coisa possível: Mas, depois, você não pode supor nada que destrua o que você primeiro supôs. Ou, se você o faz, você deve começar *de novo* [em latim, no original]. [...] Repito-o novamente: Você é livre para fazer qualquer Suposição possível: E você pode destruir uma Suposição com outra: Mas, então, você não pode reter as Conseqüências, ou qualquer parte das Conseqüências de sua primeira Suposição assim destruída (Berkeley *apud* Sautter, 2008, p. 219).

Também Peter Strawson, em *Introdução à Teoria Lógica* (1952), utilizou diversas analogias para chegar ao mesmo resultado:

Suponhamos que um homem se ponha a caminhar para um determinado lugar, mas, no meio do caminho, vira-se e regressa. Isso pode não ter sido em vão. Talvez quisesse somente um pouco de exercício. Mas, do ponto de vista de mudança de posição, é como se ele nunca tivesse começado. Semelhantemente, um homem que se contradiz pode ter sido bem sucedido no exercício de suas cordas vocais. Contudo, do ponto de vista da divulgação de informação ou da comunicação de fatos (ou falsidades), é como se nunca houvesse aberto a sua boca. Profere palavras, mas não diz nada. Ou se podia

compará-lo ao gesto de um homem que está por entregar um objeto, mas que o retém sem entregá-lo. Provoca expectativas que, em seguida, não cumpre, e este pode ter sido seu propósito. Similarmente, o propósito de um homem que se contradiz pode ser o de desconcertar. O ponto é que o propósito normal do discurso, a intenção de comunicar algo, foi frustrado pela contradição. Contradizer-se é como escrever algo para, depois, apagá-lo ou riscá-lo. A contradição se cancela e não deixa nada (Strawson *apud* Sautter, 2008, p. 219-220).

Infelizmente, ao menos de um ponto de vista formal, a “lógica proposicional” sugerida por Sautter não tem semântica matricial finita (ver SAUTTER, 2008, p. 224-225)<sup>17</sup>.

Uma solução mais satisfatória poderia ser dada em termos da parte “sadia” de um conjunto de proposições: se um conjunto de proposições é consistente, ou seja, se o conjunto de proposições é “plenamente sadio”, então suas consequências lógicas são suas consequências lógicas clássicas; e se o conjunto de proposições é inconsistente, ou seja, se o conjunto de proposições não é “plenamente sadio”, então suas consequências lógicas são as consequências lógicas clássicas do maior subconjunto “plenamente sadio” dele. Aqui destaco uma observação importante: um conjunto inconsistente de proposições pode ter mais de um subconjunto consistente, tal que esse subconjunto consistente não é subconjunto de qualquer outro subconjunto consistente, ou seja, pode haver mais de um subconjunto maximalmente consistente<sup>18</sup>. Minha sugestão é extrair as consequências lógicas clássicas do subconjunto formado pelas proposições presentes em todos os subconjuntos maximalmente consistentes; esse subconjunto poderia ser denominado, novamente recorrendo à metáfora da “saúde”, de “incontestavelmente sadio”, “sadio sob todos os aspectos”<sup>19</sup>.

---

<sup>17</sup> Ainda que não tenha semântica matricial finita, não é difícil obter um procedimento de decisão para ela baseado num procedimento de decisão para a Lógica Proposicional Clássica.

<sup>18</sup> Em termos algébricos, isso corresponde à maximalidade de tais conjuntos no reticulado de subconjuntos de um dado conjunto.

<sup>19</sup> Não é difícil de mostrar, dada a decidibilidade da Lógica Proposicional Clássica, que, caso o conjunto de proposições inconsistente seja finito, há um procedimento de decisão para essa nova “lógica”.

## Considerações finais

Vimos que:

- a) Dispomos de mecanismos eficazes para identificar excentricidades lógicas, quando elas não estão aparentes; por exemplo, os mecanismos de Quine e de Hailperin para identificar situações que não requerem um domínio do discurso não vazio.
- b) Em alguns casos, dispomos de soluções alternativas viáveis e podemos eliminar as excentricidades lógicas, se assim o quisermos, e essas soluções não contrastam em demasia da porção normal de nosso *corpus* lógico; por exemplo, as consequências clássicas da interseção de subconjuntos maximalmente consistentes é uma boa alternativa ao Princípio de Explosão.
- c) Em outros casos, a eliminação das excentricidades lógicas provavelmente não paga o preço, ou seja, o custo da eliminação seria excessivamente alto face ao benefício resultante; por exemplo, uma semântica alternativa da condicional, para o caso proposicional, distinta da semântica da condicional material, é bastante onerosa, como os projetos de lógicas relevantes revelou.

Face a isso, o melhor é seguir, no mais das vezes, a recomendação, *mutatis mutandis*, de Lya Luft mencionada como epígrafe desse texto: Vamos nos acostumar a conviver com a estranheza lógica, com a esquisitice lógica, protegendo-nos como podemos de elucubrações metafísicas desnecessárias e interpretações bizarras.

## Referências

- BARTLEY, III, W. W. **Lewis Carroll's symbolic logic**. 6. ed. New York: Clarkson Potter, 1986.
- DA COSTA, N. C. A. **Sistemas Formais Inconsistentes**. 2. ed. Curitiba: Editora UFPR, 1993 [1963].
- GARRIDO, M. **Lógica simbólica**. 6ª. reimpresión revisada. Madrid: Tecnos, 1983.
- GOLDSTEIN, L. The Barber, Russell's Paradox, Catch-22, God and More. In: PRIEST, G.; BEALL, J. C.; ARMOUR-GARB, B. **The law of non-contradiction: New philosophical essays**. Oxford: OUP, 2004.

HAILPERIN, T. Quantification theory and empty individual-domains. **The Journal of Symbolic Logic**, v. 18, n. 3, 1953, p. 197-200.

HELLER, J. **Catch-22**. New York: Simon & Schuster, 1961.

LINDEMANN, J. L.; SAUTTER, F. T. A abordagem carrolliana a paradoxos. **Pensando – Revista de Filosofia**, v. 10, n. 20, 2019, p. 91-106.

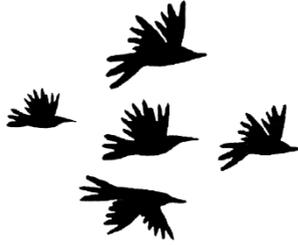
QUINE, W.V. Quantification and the empty domain. **The Journal of Symbolic Logic**, v. 19, n. 3, 1954, p. 177-9.

SÁNCHEZ VALENCIA, Victor. The algebra of logic. In: GABBAY, D; WOODS, J. **Handbook of the History of Logic**, Vol 3 The Rise of Modern Logic. Amsterdam: Elsevier, 2004.

SAUTTER, F. T. Uma expressão formal da noção kantiana pré-crítica de oposição. **Filosofia Unisinos**, v. 9, n. 3, 2008, p. 214-27.

SAUTTER, Frank Thomas. Ao modo de Saccheri. **Principia**, v. 23, n. 2, 2019, p. 145-51.

SMULLYAN, Raymond M. **What is the Name of this Book?** – The Riddle of Dracula and other logical puzzles. Englewood Cliffs: Prentice-Hall, 1978.



## Desacordos profundos e revisão da lógica: uma proposta neo-pragmatista <sup>1</sup>

*Marcos Silva* <sup>2</sup>

### Introdução

O programa intuicionista de Brouwer, como apresentado em sua histórica tese de doutorado, e também, em seu artigo “The unreliability of the logical principles” (1908), mostra uma motivação filosófica importante para a revisão de princípios lógicos e não somente para inovações de engenharia na manipulação regrada de símbolos e na introdução de uma pletera de cálculos formais. Nos sistemas intuicionistas de Heyting (1928), por exemplo, uma inferência é válida se houver uma construção que torne verdadeira a fórmula que é obtida ao aplicarmos uma regra. Nesta visão, o que os princípios da lógica deveriam preservar não deveria ser, como na lógica clássica, uma verdade independente de mentes e de atividades cognitivas, mas construtibilidade. Assim, vários princípios da lógica

---

<sup>1</sup> Este trabalho contou com o apoio do edital Universal 2018 do CNPq 430682/2018-2.

<sup>2</sup> Atualmente, é professor adjunto do Departamento de Filosofia da Universidade Federal de Pernambuco (UFPE) e tem como área de concentração a Filosofia da Lógica. É bolsista de produtividade PQ2 do CNPq e atual vice-presidente da Sociedade Brasileira de Filosofia Analítica (SBFA). Desde 2020, atua como Editor Chefe da "Perspectiva Filosófica", revista do PPGFIL da UFPE. Em 2017, ganhou o FULBRIGHT JUNIOR FACULTY MEMBER AWARD e foi pesquisador visitante na University of Pittsburgh, supervisionado por Robert Brandom.

clássica, mais notadamente o princípio do terceiro excluído, deveriam ser tomados como insuficientemente fundados.

A falta de fundamento razoável para o princípio clássico da explosão também apresenta razões importantes para a revisão de princípios lógicos, porque deveria ser problemático derivar toda sentença possível de uma contradição (Da Costa 1958, 1959, 1974). Na lógica clássica, é suficiente ter uma contradição para que isto redunde em uma relação de consequência lógica explosiva nestes sistemas. Nós deveríamos, então, distinguir trivialização de contradição. De fato, nosso problema em domínios racionais não deveria ser a existência de contradições, mas sim a ocorrência de uma relação explosiva de consequência lógica<sup>3</sup>. Nós deveríamos, por conseguinte, distinguir trivialização de contradição (Marcos, 2004; 2010). Em algumas discussões racionais ou em alguns domínios importantes de nossas vidas, quando nós lidamos com crenças e informações, por exemplo, nós, de fato, encontramos contradições e continuamos raciocinando mesmo assim, sem a demanda irreal de nos comprometermos com qualquer outra crença, especialmente as estapafúrdias (Carnielli & Rodrigues, 2019a e 2019b). Além disso, vale notar que algumas teorias na ciência poderiam ser inconsistentes sem serem triviais (Priest, Tanaka & Weber, 2015).

Estas objeções filosóficas para a natureza e o escopo da ortodoxia clássica pavimentaram o caminho para a emergência de duas lógicas alternativas à lógica clássica, respectivamente, a lógica intuicionista e paraconsistente. Estas famílias de lógicas possuem importantes implicações filosóficas e técnicas para a matemática e para a computação também.

Além disso, nós devemos tomar a corrente pluralidade de lógicas não-clássicas como uma questão filosófica séria, isto é, como algo que demanda uma explicação filosoficamente informada e nos força a reconsiderar o papel e o propósito da lógica. Duas visões extremas impediram o caminho para nos engajarmos com a importância filosófica da pluralidade de lógicas

---

<sup>3</sup> Para uma visão alternativa que defende a razoabilidade do princípio da explosão, ver Steinberger (2016).

alternativas. Por um lado, uma visão tomaria a lógica como algum tipo de investigação ou empreendimento metafísico que poderia nos oferecer ferramentas intelectuais sofisticadas para descobrirmos estruturas abstratas independentes de nossas atividades cognitivas na realidade última das coisas. Por outro lado, outra posição manteria que a lógica é basicamente um jogo vazio de conteúdo (e de significado filosófico) governado por decisões e regras arbitrárias e *ad hoc*.

Este cenário, de fato, nos oferece um problema epistemológico geral que pode ser aplicado seminalmente na discussão concernente à natureza da lógica e à própria possibilidade de revisão de seus princípios e regras mais básicas. Poder-se-ia questionar se é possível (ou como) se escolher as regras mais básicas de um sistema. Como poderíamos racionalmente justificar nossos princípios lógicos, se a própria possibilidade de justificação racional os pressupõe? De fato, uma miríade de questões epistemológicas podem ser vistas aqui: como nós poderíamos argumentar sobre um conjunto básico de regras de inferências e princípios lógicos? Qual argumento racional poderia ser usado para convencer um interlocutor litigante que um conjunto básico de regras é o correto, se todo argumento deve ser, desde o início, baseado em um conjunto de regras inferenciais já aceitas? Estas são questões concernentes à natureza da lógica e à racionalidade e se aplicam também à viabilidade de sua justificação, uma vez que se pode questionar como a razão poderia ser usada para fundamentar os princípios mais básicos da razão sem circularidade e regresso ao infinito. Existe uma maneira racional de convencer alguém de algo tão fundamental quanto um princípio da lógica?

De acordo com Bueno & Colyvan (2004), por exemplo, a emergência de lógicas não-clássicas motiva a revisão de muitas características filosoficamente atraentes da lógica e de sua centralidade para a racionalidade e para nossas vidas cotidianas. De fato, perspectivas filosóficas que enfatizam predicados especiais para a lógica vinculando-a à necessidade, universalidade, aprioricidade, eternidade, por exemplo, foram severamente desafiadas durante as últimas décadas na pesquisa

filosófica. A respeito destas dificuldades Bueno & Colyvan (2004, p. 156) afirmam:

One of the reasons that philosophy of logic is such a difficult enterprise is that, in order to advance debates in this area, we require the very thing we are studying: logic. This difficulty is especially acute when engaging in the business of theory choice in logic. After all, in order to decide between two logical theories, we need to put forward evidence and arguments for each. This evidence and these arguments will need to be assessed, and the assessment will need to be conducted in the context of some logic or other. But how do we choose this latter logic? We appear to be headed for an infinite regress.

Note que o conhecido trilema de Agripa que desafia toda tentativa de fundamentar epistemologicamente nossas justificações e crenças também pode ser aplicado neste contexto de revisão da lógica. No desafio cético de Agripa, qualquer cadeia de justificação de nossas crenças e teorias irá eventualmente enfrentar três problemas principais. Primeiramente, como apresentado na citação acima, uma vez que entremos na investigação a respeito da natureza de princípios lógicos e como justificá-los, uma cadeia de regresso ao infinito nos ameaça. Contudo, poderíamos evitar este regresso ao aceitarmos um axioma auto-evidente ou algo similar na base da hierarquia de justificação. No entanto, esta estratégia pode ser vista como arbitrária ou dogmática por um cético. Outra saída é propor que crenças e justificações deveriam ser mutuamente fundados, ou seja, deveriam fundar umas as outras, como em uma visão coerentista. Porém, ao usarmos uma cadeia de evidências e justificações que deveriam fundamentar umas as outras, um cético poderia atacar esta estratégia inegavelmente circular. Arbitrariedade na estipulação de um primeiro axioma, circularidade ou regresso ao infinito são desafios céticos encontrados em teorias sobre o fundamento do conhecimento e também ocorrem na discussão a respeito da justificação de nossos princípios lógicos.

A maior contribuição de uma abordagem neo-pragmatista<sup>4</sup> para discussões contemporâneas acerca da revisão da lógica deve ser uma teoria filosófica da inferência lógica que explique as regras e a natureza pragmática da racionalidade e da lógica ela mesma, ao darmos centralidade e desenvolvimento à seguinte hipótese de trabalho: leis lógicas deveriam ser legitimamente tomadas, a partir de uma abordagem antropológica, como um conjunto de regras que governam nossas práticas discursivas em uma esfera pública. Em outras palavras, princípios lógicos deveriam ser tomados como regras com poder normativo que constituem e corrigem nossas práticas em uma esfera discursiva pública de indivíduos engajados em trocas dinâmicas em comunidades. Assim, provas podem ser vistas como diálogos entre oponentes nos quais o aplicante introduz uma tese e o oponente tenta sistematicamente bloquear o estabelecimento da verdade da teoria, como mostrados em trabalhos de Dutilh-Novaes (2012; 2015; 2020), e Brandom (1994; 2000). Nesta perspectiva social, lógica deveria ser tomada como um conjunto de práticas argumentativas nas quais participantes podem ter diferentes objetivos, recuperando assim o significado próprio de debates da democracia grega. Assim, uma prova lógica, antes de ser concebida como uma ferramenta para estabelecer verdades eternas, pode representar um itinerário finito e sequencial de um discurso para convencer uma audiência. A verdade é estabelecida pela persuasão resultante de passos justificados por regras públicas e não pela eminente verdade transcendente de nossas práticas. Neste contexto, regras, acordos, estipulações desempenham o papel de correção objetiva do discurso, que, na abordagem realista, é garantida por verdades impessoais, independentes de nossas práticas inferenciais. Nesta visão neo-pragmatista, a abordagem realista da lógica deveria ser evitada como bloqueando qualquer revisão da lógica, prevenindo a emergência de uma

---

<sup>4</sup> Vale notar que apesar de enfatizar elementos pragmatistas como agência, uso, prática em atividades cognitivas, e ter uma postura claramente anti-metafísica a respeito da natureza da lógica, minhas influências não são os pragmatistas americanos, mas outros autores como o assim chamado segundo Wittgenstein e Robert Brandom.

pluralidade de lógicas não-clássicas com diferentes objetivos, escopos e aplicabilidades.

No que se segue, desenvolvo, na seção 1, a proposta de vermos a lógica como um empreendimento humano. Na seção 2, discuto a ideia de desacordos profundos no horizonte da revisão da lógica e defendo, na seção 3, que a lógica seja composta de *hinge propositions*. Concluo defendendo algumas consequências tiradas deste quadro conceitual neo-pragmatista aplicado à natureza da lógica e à possibilidade de sua revisão.

### **1) Lógica como um empreendimento humano**

Uma perspectiva neo-pragmatista para fundamentar a racionalidade, em geral, e a natureza da necessidade lógica, em particular, deve permitir a revisão da lógica. É importante enfatizar que um paradigma filosófico está sendo desafiado aqui. O representationalismo está sendo rejeitado. A racionalidade não deveria ser pensada em termos de atributos de uma entidade individual, isolada e imaterial consumindo e manipulando representações, mas como uma capacidade ou habilidade especial de alguns animais de se engajarem deontologicamente, isto é, em termos de proibições e autorizações em trocas permanentes com o meio onde estão inseridos e com outros animais com a mesma capacidade. Neste pano de fundo neo-pragmatista, a discussão semântica acerca do significado de nossos conteúdos mentais e de alguns signos em nossa linguagem pode e deveria ser desmistificada.

Se nós adotarmos princípios lógicos como regras regulando nossas atividades discursivas, determinando regras e critérios para avaliar o que é legítimo e o que é ilegítimo em nossas trocas cotidianas, nós devemos notar que nós estamos lidando, em nossa perspectiva, com noções deontológicas, como autorizações e proibições que permitem, restringem ou guiam nosso campo de ação ou espaço de manobra (*Spielraum*). Como resultado, é crucial observar que regras determinam normativamente os critérios pelos quais nós julgamos a qualidade de nossas ações e descrições.

Assim, regras, elas mesmas, não podem ser verdadeiras nem falsas, uma vez que elas fixam os critérios para nossa avaliação de alguma coisa como verdadeira ou falsa. Contudo, note que sistemas de referência, que determinam regras e critérios de avaliação, poderiam ser diferentes e podem mudar ao longo do tempo de acordo com pressões pragmáticas e ambientais.

Podemos combinar ideias encontradas em autores como, Wittgenstein (1969), Brandom (1994; 2000) e Dutilh Novaes (2015; 2020). É importante tomar a lógica com um empreendimento humano. A lógica formal, raciocínio a partir de princípios gerais, deveria ter a sua normatividade, no fim, baseada em nossas práticas discursivas atuais e em nossos raciocínios locais tomados como corretos. O problema da revisão da lógica, primariamente, não deveria estar concernido nem com a lógica formal nem com a aplicação da lógica às ciências naturais, mas com nossos raciocínios cotidianos.

O lugar primário de normatividade e de significado deveria ser nossa vida diária, nossas inferências cotidianas regulares (Dutilh Novaes & Veluwenkamp, 2017; Oza, 2020). Aqui, é importante enfatizar a dimensão social da lógica, a absorção de regras por observação e instrução e a possibilidade de correção mútua em práticas regradas comunais, que pressupõem treinamento pela transmissão, inculcação e imersão em uma comunidade humana.

A analogia a ser feita aqui, concernente à relação entre lógica clássica e lógicas alternativas, deve ser a de um modelo de regulação que não se compatibiliza com as práticas atuais. A lógica clássica, como modelo único de regulação, de normatividade, em relação a práticas inferenciais locais não é coerente com estas mesmas práticas. Desta forma, aquela deveria ser revista a partir destas. A aproximação da lógica a ser feita em relação às ciências não é com as ciências naturais, mas com ciências normativas como as jurídicas, políticas e éticas. Quando um sistema geral de regulação não cumpre seu papel normativo, ele deve ser revisto em face das práticas atuais.

## 2) Desacordos profundos e revisão da lógica

Nas últimas décadas, houve interesse crescente em como usar filosofia para pensar a respeito da natureza de desacordos profundos, de rupturas comunicacionais provindas deles e de como lidar com estes fenômenos. Fogelin (1985), por exemplo, introduz a seminal noção de *Deep Disagreement* para conceitualizar rupturas comunicacionais e desafios em dilemas éticos. Por um lado, nós temos desacordos que podem ser chamados de normais, uma vez que são trocas argumentativas nas quais os pares inicialmente entram em desacordo sobre a verdade de uma proposição qualquer, por exemplo, sobre as realizações de um time de futebol, mas, geralmente, estão de acordo sobre o que constituiria evidência a favor ou contra a afirmação controversa (Martin, 2019). De fato, o discurso racional terá provavelmente mais sucesso se existir uma boa quantidade de confiança e boa vontade entre os participantes (Davson-Galle, 1992). Com efeito, deve haver neste cenário alguma expectativa mútua entre os pares de que os recursos epistêmicos a serem trocadas são valiosos e confiáveis (Feldman, 2005). Além disso, a probabilidade de argumentação frutífera cresce se os pares adotarem regras, comportamentos e atitudes esperadas específicas (Adams, 2005).

Entretanto, por outro lado, Fogelin (1985, p. 8) também afirma que:

[...] we get a very different sort of disagreement when it proceeds from a clash in underlying principles. Under these circumstances, the parties may be unbiased, free of prejudice, consistent, coherent, precise and rigorous, yet still disagree. And disagree profoundly, not just marginally. Now when I speak about underlying principles, I am thinking about what others (Putnam) have called framework propositions or what Wittgenstein was inclined to call rules. We get a deep disagreement when the argument is generated by a clash of framework propositions.

A tese de Fogelin é a de que desacordos profundos estão além da resolução racional, porque a falha em possuir compromissos

compartilhados em desacordos profundos impediria a solução do dissenso. Fogelin (1985, p. 5) adverte que enquanto “the language of argument may persist... it becomes pointless since it makes an appeal to something that does not exist; a shared background of beliefs”. De fato, se pode perguntar que tipo de procedimentos racionais poderiam ser usados para a resolução do dissenso quando desacordos profundos, de fato, surgem. A resposta de Fogelin (1985, p. 9) é: “The drift of this discussion leads to the answer NONE”. Fogelin (1985, p. 4-5) é rigoroso em manter a força de sua tese: “not the weak claim that in such contexts arguments cannot be settled. It is the stronger claim that the conditions for argument do not exist”. Aqui, a observação (bem) negativa de Fogelin está chamando nossa atenção para a lógica informal e nossas práticas discursivas cotidianas ao invés de ferramentas da lógica formal.

A avaliação de Fogelin pode ser vista ao menos tão negativa quanto a tese que poderíamos chamar de incomensurabilidade *a la* Quine, uma vez que desafia a própria possibilidade da revisão da lógica e de uma genuína rivalidade entre lógicas. Quine (1970) mantém que lógicos rivais estão falando de coisas diferentes. Ou em outras palavras, eles aparentam estar falando um com outro, enquanto, de fato, eles estão usando linguagens diferentes. O argumento quineano se segue desta maneira: Se uma lógica diferente da clássica considera inválido uma inferência válida, então outro significado estaria sendo atribuído para operadores lógicos. Mas se lógica clássica e uma lógica alternativa atribuem significados diferentes para operadores lógicos, elas estão falando de coisas diferentes. Assim vemos o sentido do slogan “change of logic, change of subject”. A conclusão seria que se o assunto mudou, não é mais lógica, assim uma lógica não-clássica não deveria ser chamada de lógica, mas, sim, de alguma coisa diferente (1970, p. 75-80). Assim, por exemplo, um lógico intuicionista e outro clássico podem aparentar estar discutindo sobre qual seria a negação legítima em sistemas formais, mas, no entanto, não veriam que uma negação intuicionista é algo tão significativamente diferente de uma clássica que a discussão seria inútil. O resultado negativo desta apreciação

é que, se não houver nada em comum no conflito, não há conflito real, e, por conseguinte, não há a possibilidade real de dissenso e de revisão.

Contudo, não precisamos ser tão céticos acerca de desacordos profundos em lógica, se nós adotarmos o que foi introduzido acima como uma abordagem neo-pragmatista na lógica. Uma vez que a ênfase da nossa proposta não está em cálculos matemáticos “monológicos”, mas sim no nível antropológico e multi-agente, isto é, social em essencial, de nossas práticas discursivas, a tradicional abordagem da lógica muito formalista e abstrata deveria ser substituída pela investigação de práticas discursivas reais usadas em línguas naturais, uma vez que linguagem é vista como uma prática social e não como um objeto estritamente matemático (Brandom, 1994; 2000; Dutilh Novaes, 2015; 2020).

Assim, semânticas devem ser baseadas em termos do significado na medida em que este é fixado pelo uso de agentes racionais. O paradigma de coisas sendo referenciadas a partir de simples expressões, isto, é o paradigma representacional, deve ser substituído pelo paradigma que examina o que nós fazemos em práticas para estabelecer esta associação. Há, pois, uma importante mudança da noção de representação alcançando uma realidade independente para o debate acerca de acordos e práticas no uso de algumas ferramentas e construções abstratas. A discussão clássica acerca de crenças, conteúdo mental e do “*know that*” teórico é substituída pela investigação de habilidades e competências de indivíduos a serem treinados em ambientes educacionais e do prático “*know how*”.

Com efeito, a lógica formal não captura a complexidade de práticas argumentativas ordinárias muito simples e muito menos a complexidade em desacordos profundos onde nós temos rupturas comunicacionais. De acordo com neo-pragmatistas, como Fogelin (1985) e Brandom (1994; 2000), nós devemos começar com maneiras não-formais de raciocínio para entender a origem da normatividade da sistemas formais. Novaes (2015; 2020) também defende que a característica normativa da lógica primariamente deveria ser encontrada no raciocínio enquanto tomado como um processo social e não como um processo mental ou uma

investigação metafísica, uma vez que nossa atenção deveria ser dada “to dialogical interactions between two or more agents.” (2015, p. 588).

### 3) Sobre lógica como *Hinge Propositions*

Ao combinarmos outras importantes ideias encontradas em Wittgenstein (1969), Dutilh Novaes (2013; 2015; 2016), Fogelin (1985) e Pereira (2006), nós podemos motivar uma abordagem neo-pragmatista para atacar o problema de desacordos profundos sobre a revisão da lógica ao oferecermos uma proposta não-fundacionista para a natureza da lógica. A ênfase para fazer a revisão da lógica possível, sem identificar a atividade lógica com uma atividade metafísica, não deveria ser posta na investigação de alguma lei profunda ou escondida da natureza ou da realidade, mas sim em um tipo de equilíbrio de práticas humanas baseadas em interações sociais e linguísticas complexas, isto é, sobre o longo e complexo histórico da associação dinâmica entre formas de vidas e princípios da lógica como *hinge propositions*. É uma falsa dicotomia pensar que a natureza da lógica tem seus fundamentos ou na arbitrariedade das regras de manipulação de signos vazios de significado ou ser fundada na realidade última das coisas. A fundação deveria ser social e estável ou regular o suficiente. Ela deveria ter poder normativo, isto é, ela deveria ser usada por nós para corrigirmos casos deviantes.

Neste sentido, podemos ter boas razões para introduzir a assim-chamada *Hinge epistemology* (Coliva, 2015; Moyal-Sharrock, 2016; Pritchard, 2012; 2016; Ranalli, 2018a; 2018b) no contexto da revisão da lógica. Moore (1925; 1939) usa “eu sei” conectado com ou prefixando diversas proposições empíricas diferentes para desafiar o idealismo e o ceticismo. Alguns exemplos destas proposições são: “I am a human being”, “there are other human beings”, “I have a brain”, “I’ve never been out of our planet”, “I’ve never been on the moon”, “I have two hands”, “here it is a hand”, “Every human being has a brain”. Wittgenstein, em seu *On Certainty* (1969, OC) questiona como aprendemos estas “verdades”. A

verdade de proposições de Moore é certamente não *a priori* e não é garantida por verificação; afinal, poucos de nós devem ter visto um cérebro humano de fato ao longo de suas vidas. Elas não parecem se basear em uma indução formal tampouco, porque elas não são baseadas em uma generalização comum. Se forem baseadas em uma generalização, digamos, de um biólogo, seria uma generalização fundada em muito poucos espécimes; afinal, quantas pessoas e seus cérebros o biólogo deve ter examinado de fato?

Duas propriedades destas *hinge propositions* são imediatamente aparentes para Wittgenstein. Primeiro, elas são requeridas para investigar a verdade de outras proposições, e segundo, elas devem ser isentas de dúvida. Nós não usualmente articulamos estas “verdades” e não aceitamos de fato sua possível falsidade lógica. Central para todas elas é a ideia que *hinges* desempenham um peculiar papel em nossas práticas discursivas, um papel semelhante a uma regra (OC 95), que as faz, em última palavra, imunes a assaltos céticos ao tirá-las o estatuto de conhecimento. Por conseguinte, proposições como “Here’s my hand” (em cenários como os de Moore), “Nobody has ever been on the Moon” (antes de 1949), “The Earth has existed for a very long time”, “My name is Marcos Silva” e “There are physical objects” deveriam ser tomadas como regras de “evidential significance” (Wright, 1985), ou “norms of description” (OC 167, 321). Em outras palavras, justificação e conhecimento são possíveis, assim como a dúvida, somente dentro dos limites determinados ao se tomar as *hinge* como garantidas. Elas desempenham o papel regulativo, ou normativo, para a avaliação a qualidade de outras asserções e demais práticas racionais. Assim, como nós podemos afirmar, elas são elementos constitutivos de nossa racionalidade epistêmica.

Neste pano de fundo, nossa hipótese de trabalho é dividida em três partes, a saber: i) os princípios lógicos deveriam ser tratados como *hinge propositions* (mas nem todas as *hinge propositions* são princípios lógicos), ii) elas são determinadas pela nossa educação em uma *Weltbild*; iii) os princípios lógicos poderiam ter sido diferentes e podem mudar.

As *hinge propositions* não são metafísicas; não redundam em conhecimento positivo da essência da realidade última das coisas. Elas não são descrições tampouco. Elas não possuem uma forma lógica comum (de fato, elas possuem, com frequência, a forma de proposições empíricas) e elas não possuem um conteúdo semântico comum tampouco. Elas não são dubitáveis em função do papel peculiar que desempenham em nossa forma de vida. Este papel peculiar as tornam lógica. E este papel pode ser modificado. Nós somos educados por elas e através delas. Como em conflitos radicais, na lógica nós também temos que iniciar litigantes em novos procedimentos, interesses e visões; ao modificarmos suas *hinge propositions*, modificamos sua *Weltbild*. A ideia básica é que nós temos uma fundação, mas sem fundacionismo. Segundo esta abordagem neo-pragmatista, em alguns cenários e situações, o oposto nunca é considerado, porque alguma coisa deve permanecer sólida para que nossas atividades teóricas e práticas sejam possíveis. Aqui não se trata da questão de um conteúdo proposicional especial por representar fatos especiais da realidade, mas sim de nossas maneiras de agir e julgar tirarem de consideração e de dúvida algumas coisas.

Nós podemos motivar a transição de *hinge epistemology* para a lógica, ao enfatizarmos a analogia entre *hinge* e princípios lógicos. Ambos são requeridos para investigar o sentido e a verdade de outras proposições. Eles são isentos de dúvida, porque algo deve permanecer firme (caso contrário, nós "perderíamos o pé", minariamos ou paralisariamos nossas investigações). Em sistemas de crenças, como em sistemas lógicos, algumas coisas devem permanecer inabaláveis, como absolutamente sólidas e seguras. Note que *hinge propositions* são apresentados na literatura como "óbvias", assim como princípios lógicos, com frequência, são também apresentados. Rejeição ou recusa de *hinge propositions* e de princípios lógicos são, com frequência, vistas como imediatamente insanas, não apenas falsas, uma vez que nós não abriríamos mão delas porque elas pertencem a coisas das quais estamos profundamente

convencidos. Como Wittgenstein afirma: “the reasonable man does not have certain doubts” (OC 220).

A discussão sobre normatividade partindo da lógica clássica e do monismo não parece representar grandes avanços filosóficos. Ora, dentro do horizonte de que há somente uma lógica verdadeira ou legítima e que esta é a lógica clássica, parece ser fácil determinar a normatividade da lógica, porque qualquer desvio da lógica clássica seria ou raciocinar incorretamente ou mesmo não raciocinar (Steinberger 2017a; 2017b). Contudo, é importante observarmos três pontos: i) a diversidade contemporânea de lógicas alternativas com aplicações muito seminais em áreas diversas e inalcançáveis pela ortodoxia da lógica clássica, ii) a abordagem da relação de consequência lógica a partir da teoria da prova onde o significado do vocabulário lógico é definido em termos de regras de introdução e eliminação, e iii) como vimos, a filosofia do segundo Wittgenstein, ao enfatizar práticas humanas como centrais para entender a normatividade da linguagem e não a referência a entidades independentes de toda e qualquer atividade cognitiva humana.

Com frequência, a normatividade da lógica é deslocada para a normatividade da epistemologia em função da aproximação da lógica com ciências naturais (Harman, 1984; 1986). A tese é que a física e a lógica só se distinguiriam em escopo, e não em tipo. Acredito que associar a natureza da lógica com a natureza de ciências empíricas está relacionado, em última instância, a uma falsa dicotomia, a saber, entre: prescrições e descrições. O fenômeno da normatividade é muito mais abrangente e complexo que o fenômeno da prescrição. Sentenças prescritivas não cobrem o largo espectro do significado filosófico de elementos normativos. Toda prescrição é normativa, mas nem toda a normatividade redunde em alguma forma direta ou indireta de prescrição. Se pensarmos em construtivistas, como os intuicionistas, os operadores lógicos devem ser tomados como normas a partir da identificação do significado de operadores com regras de introdução e eliminação. Se pensarmos na tradição kantiana que influi, por exemplo, em algumas discussões em

Brandom (1994; 2000) e em Wittgenstein (1986), conceitos, qualquer um, são normas, são regras, uma vez que aprender um conceito é aprender sob quais condições devemos aplicá-lo legitimamente. Dominar um conceito para um neo-pragmatista é dominar o uso significativo de uma palavra em nossas línguas naturais.

A abordagem naturalista da lógica que associa lógica a ciências naturais parece pressupor que nós operamos conceitos, depois fazemos ciência e depois temos normatividade. Contudo, vale notar que nós temos normatividade sempre que usamos conceitos. Ao usarmos qualquer conceito, já estamos em um âmbito normativo de permissões, proibições, autorizações, obediência a regras etc. Desta forma, a discussão contemporânea a respeito do anti-exceptionalismo ao associar lógica a uma ciência empírica parece desencaminhadora, porque conflui descriptivismo e realismo lógico (Mcsweeney, 2018). A lógica, como qualquer outra ciência, seria responsável por descrições verdadeiras de um domínio muito geral de coisas abstratas.

Não é controverso que a lógica booleana pode ser aplicada para a compreensão do comportamento de circuitos elétricos. Contudo, para a filosofia, é importante perguntar o que significa isto. Qual seria a interpretação filosófica mais adequada deste fenômeno de aplicação? Aplicar convenientemente a lógica booleano para entender o comportamento de circuitos elétricos significa que a lógica descreve corretamente circuitos elétricos? Analogamente, se um jogo de xadrez puder ser aplicado a uma batalha de fato, isto significa que o xadrez descreve verdadeiramente a batalha? Ou mesmo se a constituição do Brasil puder ser aplicada sistematicamente a um outro país, isto significa que ela descreve verdadeiramente a política deste país? Um jogo de xadrez e a constituição de um país são conjuntos de regras, fixam significados, procedimentos, autorizações e proibições de atividades sociais colaborativas e não precisam descrever nada na realidade, assim como a lógica booleana.

Um dos motes filosóficos de Brandom é que nós usamos princípios lógicos para corrigir, regular nossas ações, percepções, interações, teorias, informações etc. E não, primariamente, descrever coisas. Ele usa, em sua abordagem neo-pragmatista da lógica, alguns tópicos do anti-realismo lógico de acordo com o qual vocabulário lógico não se relaciona com nenhum estado de coisa particular do mundo, mas com os nossos critérios e normas para avaliar descrições e ações no mundo. Nesta visão, nós não estamos falando sobre fatos e verdade, mas sobre os nossos critérios para avaliar fatos e verdade. Assim, os princípios lógicos não precisam representar alguma coisa na realidade. Brandom (1994; 2000) também introduz a tese importante do expressivismo lógico de acordo com a qual regras e vocabulário lógico expressam, mostram, tornam explícitos, possíveis formas de representar o nosso mundo. Sistemas lógicos expressam alguns comprometimentos e normas em discussões e práticas racionais cotidianas. Sustentando seu antirealismo e expressivismo, Brandom defende o inferencialismo semântico que mantém que entendimento, significado e cognição devem ser associadas com a articulação de inferências materiais e não com referência a objetos especiais independentes da linguagem. Nossas práticas regradas, as formas que agimos no mundo são articuladas inferencialmente e podem ser publicamente testadas e controladas.

Tendo em mente o expressivismo lógico de Brandom, é agora mais natural aplicar a *hinge epistemology* ao problema da revisão da lógica. De acordo com o projeto neo-pragmatista proposto aqui, a distinção entre proposições empíricas e proposições das lógicas não é nem baseada no seu conteúdo, nem em sua forma lógica, mas sim no seu papel ou função em nossa *Weltbild*, isto é, em nosso entendimento teórico do mundo e de como operá-lo. Proposições empíricas podem também, algumas vezes, desempenhar o papel de normas, quando elas são usadas, por exemplo, para corrigir nossa percepção e guiar nossas ações e investigações.

## Conclusão

Nossa abordagem é uma proposta anti-realista. Ao enfatizarmos a natureza normativa da lógica, defendemos que princípios lógicos deveriam ser tomados como mais próximos de princípios éticos que de leis científicas, uma vez que eles não precisam descrever qualquer fato na realidade, mas normativamente guiam nossas decisões e comportamentos. Em nossa proposta, agentes racionais não precisam estabelecer fatos, mas eles estabelecem os critérios que são usados para descrever fatos e avaliar a qualidade de nossas descrições.

Nós não podemos argumentar a favor ou contra escolhas para regras básicas como nós podemos argumentar para ou contra um passo em um raciocínio dentro de um sistema definido; nós só podemos aceitar ou rejeitar estas regras básicas; ademais, nos deveríamos, sim, ser instruídos ou educados para usar algumas regras. De acordo, como nossas regras inferenciais poderiam ser diferentes, nossa abordagem neo-pragmatista foi usada para fundamentar a possibilidade de uma pluralidade de lógicas não-clássicas e como nós podemos ter muitos sistemas lógicos rivais sem a necessidade de tomar algum como sendo o sistema lógico correto.

O problema da objetividade é talvez o desafio conceitual mais sério ao enfrentarmos qualquer tentativa de fundamentar as propriedades governando o uso de conceitos em práticas sociais. Longe de impedir a possibilidade da objetividade conceitual, entender o caráter essencialmente social da prática discursiva, na qual normas conceituais são implícitas, é o que torna tal objetividade inteligível. Entender racionalidade é entender a possibilidade de correção e de objetividade de nossos discursos, uma vez que garantem a normatividade de regras públicas que seguem, guiam e constituem nossas práticas.

A proposta aqui defendida, concernente ao problema da justificação e normatividade na revisão da lógica, desenvolveu alguns tópicos apresentados no "Sobre a Certeza" de Wittgenstein, oferecendo uma visão neo-pragmatista sobre a pluralidade de lógicas não-clássicas. Apliquei a

*hinge epistemology* de Wittgenstein e seu anti-realismo à discussão sobre a normatividade em meio ao desacordo profundo entre lógicas rivais. Assim, rejeitei a abordagem realista acerca da natureza da lógica e propus uma visão na qual a natureza de princípios lógicos está relacionada com *hinge propositions*. Assim, uma explicação para a normatividade da lógica no ambiente de desafio e revisão de seus princípios foi apresentada enfatizando elementos neo-pragmatistas, uma vez que o modelo de regulação clássico não parece se compatibilizar mais com nossas práticas inferenciais atuais.

## Referências

- ADAMS, D. M. Knowing when disagreements are deep. **Informal Log**, 25, p. 65-77, 2005.
- BUENO, O. & COLYVAN, M. Logical non-apriorism and the law of non-contradiction. In: Graham Priest, Jc Beall & Bradley P. Armour-Garb (eds.). **The law of non-contradiction: new philosophical essays**. Oxford University Press, 2004, p. 156-75.
- BRANDOM, R. **Making it explicit: reasoning, representing, and discursive commitment**. Harvard University Press, 1994.
- \_\_\_\_\_. **Articulating reasons**. Cambridge, MA: Harvard University Press, 2000.
- BROUWER, L. E. J. **Over de Grondslagen der Wiskunde** (On the Foundations of Mathematics). Ph.D. thesis, Universiteit van Amsterdam. English translation In: Brouwer, 1975, p. 11-101 [1907].
- \_\_\_\_\_. **De onbetrouwbaarheid der logische principes** (The Unreliability of the Logical Principles). Tijdschrift voor Wijsbegeerte, 2, p. 152-8. English translation in Van Atten and Sundholm 2017. An older English translation is In: Brouwer, 1975, p. 107-11. doi:10.1016/B978-0-7204-2076-0.50009-X [1908].
- CARNIELLI, W. & RODRIGUES, A. On epistemic and ontological interpretations of intuitionistic and paraconsistent paradigms. **Logic Journal of the IGPL**, 2019a. Doi: 10.1093/jigpal/jzz041.

- \_\_\_\_\_. Inferential semantics, paraconsistency, and preservation of evidence. In: BASKENT, C. & FERGUSON, T. M. (eds). **Graham Priest on dialetheism and paraconsistency**. Springer, 2019b.
- COLIVA, A. **Extended rationality**: a hinge epistemology. Palgrave Macmillan, Basingstoke, 2015.
- COLIVA, A. & MOYAL-SHARROCK, D. (eds.). **Hinge epistemology**. Brill, Leiden, 2016.
- DA COSTA, N. C. A. Nota sobre o conceito de contradição. **Anais da Sociedade Paranaensed de Matemática**, 1, p. 6-8, 1958.
- \_\_\_\_\_. Observações sobre o conceito de existência em matemática. **Anais da Sociedade Paranaense de Matemática**, 2, p. 16-9, 1959.
- \_\_\_\_\_. On the theory of inconsistent formal systems. **Notre Dame J Form L**, 15, p. 497-510, 1974.
- DAVSON-GALLE, P. Arguing, arguments, and deep disagreements. **Informal Log**, 14, p. 147-56, 1992.
- DUTILH NOVAES, C. A dialogical, multi-agent account of the normativity of logic. **Dialectica**, 69, p. 587-609, 2015.
- \_\_\_\_\_. **The dialogical roots of deduction**. Cambridge: Cambridge University Press, forthcoming, 2020.
- DUTILH NOVAES, C. & VELUWENKAMP, H. Reasoning Biases, Non-Monotonic Logics and Belief Revision. **Theoria**, 83, p. 29-52, 2017. Doi:10.1111/theo.12108.
- FELDMAN, R. Deep disagreement, rational resolutions, and critical thinking. **Informal Log**, 25, p. 13-23, 2005.
- FOGELIN, R. The Logic of Deep Disagreements. **Informal Log**, 7, p. 1-8, 1985.
- HARMAN, G. Logic and Reasoning?, **Synthese**, 60, 1, p. 107-27, 1984. Doi:10.1007/BF0048562
- \_\_\_\_\_. **Change in view**: principles of reasoning. Cambridge, MA: MIT Press, 1986.

HEYTING, A. **Prize essay on the formalization of intuitionistic logic**. Expanded and revised version published as Heyting 1930, Heyting 1930a, Heyting 1930b [1928].

\_\_\_\_\_. Die formalen Regeln der intuitionistischen Logik I. **Sitzungsberichte der Preussischen Akademie der Wissenschaften**, p. 42-56. English translation in Mancosu, 1998, p. 311-27 [1930].

\_\_\_\_\_. Die formalen Regeln der intuitionistischen Logik II. **Sitzungsberichte der Preussischen Akademie der Wissenschaften**, p. 57-71 [1930a].

\_\_\_\_\_. Die formalen Regeln der intuitionistischen Logik III. **Sitzungsberichte der Preussischen Akademie der Wissenschaften**, p. 158-69 [1930b].

MARCOS, J. **Logics of Formal Inconsistency**. Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Filosofia e Ciências Humanas, 2004.

\_\_\_\_\_. 2010. Wittgenstein & Paraconsistência. **Principia**, 14, 1, p. 135-73.

MARTIN, B. Searching for Deep Disagreement in Logic: The Case of Dialetheism. **Topoi**, 2019. Doi: <https://doi.org/10.1007/s11245-019-09639-4>

MCSWEENEY, M. Realismo lógico e a metafísica da lógica. Trad. Evelyn Erickson. **Perspectiva Filosófica**, vol. 45, n. 1, p. 122-40, 2018.

MOYAL-SHARROCK, D. The animal in epistemology: Wittgenstein's enactivist solution to the problem of regress. In: COLIVA, A. & MOYAL-SHARROCK, D. (eds). **Hinge epistemology**. Leiden: Brill, 2016, p. 24-47.

MOORE, G. E. A Defence of Common Sense and Proof of an External World. In: BALDWIN, Thomas (ed.). **Selected writings**. London: Routledge, 1993.

OZA, M. The value of thinking and the normativity of logic. forthcoming. **Philosophers' Imprint**.

PEREIRA, L. C. Breves considerações sobre o nilismo e o revisionismo na lógica. **O que nos faz pensar** n. 20, p. 91-9, 2006.

PRIEST, G.; TANAKA, K.; WEBER, Z. Paraconsistent Logic. In: E. N. Zalta (ed.). **The Stanford Encyclopedia of Philosophy**. Spring Edition, 2015. Disponível em: <https://plato.stanford.edu/entries/logic-paraconsistent/>

PRITCHARD, D. Wittgenstein and the groundlessness of our believing. **Synthese**, 189, p. 255-72, 2012.

\_\_\_\_\_. **Epistemic angst: radical skepticism and the groundlessness of our believing**. Princeton: Princeton University Press, 2016.

QUINE, W. **Philosophy of Logic**. Harvard University Press, 1986 [1970].

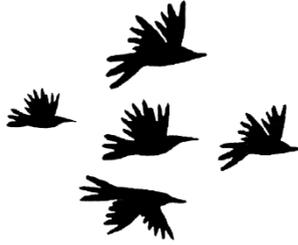
RANALLI, C. Deep disagreement and hinge epistemology. **Synthese**, 2018a. Doi: <https://doi.org/10.1007/s11229-018-01956-2>

\_\_\_\_\_. What is Deep Disagreement? **Topoi**, 2018b. Doi: <https://doi.org/10.1007/s11245-018-9600-2>.

STEINBERGER, F. Explosion and the Normativity of Logic. **Mind**, 125, p. 383-419, 2016. Doi: 10.1093/mind/fzv203

WITTGENSTEIN, L. **Über Gewißheit/On certainty**. Ed. Anscombe G. E. M. & von Wright G. H.; trans. Paul, D. & Anscombe, G. E. M. Blackwell, Oxford, 1969.

WRIGHT, C. Facts and certainty. **Proceedings of the British Academy**, 71, p. 429-72, 1985.



## A teoria supervalorativista da vagueza e o problema da precisão

*Sagid Salles*<sup>1</sup>

### Introdução

O fenômeno da vagueza é quase onipresente na linguagem natural. Por um lado, a vagueza abrange variadas categorias lógicas, incluindo predicados, termos singulares e quantificadores. Por outro, é argumentável que a maioria das expressões em cada uma destas categorias, principalmente termos singulares e predicados, é vaga. Isto não seria um problema, não tivesse o fenômeno da vagueza relacionado ao paradoxo sorites, que supostamente mostra a incoerência destas expressões. O Supervalorativismo fornece uma explicação do fenômeno, acompanhado de uma solução do paradoxo, que é interessante por pelo menos duas razões. Primeiro, é argumentável que esta teoria é a que melhor acomoda certa imagem intuitiva da vagueza, conhecida como *indecisão semântica*. Segundo, ela é relativamente conservadora com respeito à lógica clássica. Ao contrário de algumas rivais, como a Teoria Trivalente e o Gradualismo, o Supervalorativismo nos permite manter

---

<sup>1</sup> Graduado em filosofia pela Universidade Federal de Ouro Preto. Mestre e Doutor pelo Programa de Pós-Graduação em Lógica e Metafísica da Universidade Federal do Rio de Janeiro, com sanduiche realizado na University of Miami.

tanto o terceiro excluído como a não contradição. A despeito de suas aparentes vantagens, no entanto, a teoria carrega alguns sérios problemas. Com foco na categoria dos predicados vagos, meu objetivo neste artigo é mostrar que o Supervalorativismo falha em satisfazer pelo menos um de três critérios de adequação para uma teoria ideal da vagueza: o critério da precisão. O artigo é dividido em quatro seções. Na primeira, apresento o paradoxo sorites e três critérios de adequação para uma teoria ideal da vagueza: o critério do sorites, o critério da coerência e o critério da precisão. Na segunda e terceira, apresento o Supervalorativismo e sua solução para o paradoxo sorites. Por fim, argumento que, mesmo que a teoria satisfaça os dois primeiros critérios de adequação, ela certamente não satisfaz o terceiro<sup>2</sup>.

## 1. Vagueza, sorites e os três critérios de adequação

A característica mais saliente dos predicados vagos é que eles são aparentemente suscetíveis ao paradoxo sorites<sup>3</sup>. Este paradoxo, descoberto por Eubulides de Mileto no século IV a. c., atualmente, admite muitas versões (HYDE, 2011). No que segue, apresento uma das mais comuns: a versão quantificada.

Seja “C” o predicado “careca”, “a” um nome arbitrário de pessoas e deixe os números na direita inferior indicarem o número de cabelos que a tem na cabeça. Com isto em mente, podemos formular o paradoxo, em cada direção, como segue.

---

<sup>2</sup> Este capítulo é o segundo de uma série de textos, de estrutura similar, em que mostro como cada teoria da vagueza falha em satisfazer o critério da precisão. O primeiro da série é Salles (2019), no qual discuto o caso da Teoria Trivalente da Vagueza. Os trabalhos são escritos de forma que possam ser lidos independentemente. Por isto, há algumas repetições que poderiam ter sido evitadas se eles fossem todos parte de um único texto. Por exemplo, a seção 1 – em que apresento o paradoxo sorites – é inevitavelmente similar em cada um. De fato, aqueles que já leram a publicação de 2019 podem simplesmente saltar a primeira seção deste capítulo.

<sup>3</sup> É difícil recusar que exista uma relação entre o fenômeno da vagueza e a aparente suscetibilidade dos predicados vagos ao paradoxo sorites, mas isto não significa que todo predicado vago é aparentemente suscetível ao paradoxo sorites. Soames (1999, p. 217), Weatherson (2009, p. 80), Shapiro (2006, p. 4) e outros rejeitam que ser aparentemente suscetível ao paradoxo sorites é uma condição necessária para ser vago. Acredito que eles estão certos quanto a isto. Para nossos propósitos neste artigo, contudo, podemos ignorar estas sutilezas.

(SC)	(¬SC)
(1) $Ca_0$	(1) $\neg Ca_{10.000}$
(2) $\forall n (Ca_n \rightarrow Ca_{n+1})$	(2) $\forall n (\neg Ca_{n+1} \rightarrow \neg Ca_n)$
_____	_____
(3) $Ca_{10.000}$	(3) $\neg Ca_0$

Cada um dos argumentos pareados acima parte de uma suposição obviamente verdadeira para uma conclusão aparentemente falsa. (SC) parte da suposição de que uma pessoa com o fios de cabelo é careca para a conclusão de que uma pessoa com 10.000 fios também é careca. (¬SC) parte da suposição de que uma pessoa com 10.000 fios não é careca, para a conclusão de que uma pessoa com o fios também não é. Em cada caso, a conclusão se segue por repetidas aplicações de Instanciação Universal e *Modus Ponens*. Tomados em conjunto, estes argumentos nos permitem mostrar, para cada indivíduo que considerarmos, que ele é careca e não careca ao mesmo tempo. Portanto, o predicado “careca” é incoerente. O mesmo paradoxo pode ser formulado para muitos outros predicados da linguagem natural, como “alto”, “baixo”, “rico”, “pobre”, “vivo”, “morto”, “rápido”, “lento”, “monte” etc. Isto significa que cada um destes predicados é incoerente, e nossa linguagem nos deixa com contradições por toda parte. Esta conclusão é certamente inaceitável.

Se não podemos aceitar a conclusão do paradoxo, resta-nos apontar o que há de errado com ele. Podemos seguir por dois caminhos: rejeitar alguma das regras de inferência empregadas na derivação ou rejeitar alguma das premissas. A princípio, a ideia de rejeitar a Instanciação Universal ou o *Modus Ponens* é indigesta, assim como a ideia de rejeitar a premissa (1) de cada argumento acima. O caminho mais promissor seria rejeitar (2), a premissa quantificada do paradoxo. Infelizmente, contudo, uma rápida análise nos mostra que este caminho não é tão melhor do que os outros. Suponha que a premissa quantificada de (SC) é falsa, de tal modo que o seguinte seja o caso:  $\neg \forall n (Ca_n \rightarrow Ca_{n+1})$ . Isto é logicamente equivalente a  $\exists n (Ca_n \wedge \neg Ca_{n+1})$ . Por sua vez, isto significa que existe um número  $n$  tal que uma pessoa com  $n$  fios de cabelo na cabeça é careca e uma pessoa com  $n+1$  fios não é careca. Obviamente, a negação da premissa

(2) de ( $\neg$ SC) teria a mesma consequência. Em cada caso, a consequência é que o predicado “careca” corta o mundo de forma *cirúrgica*, estabelecendo uma fronteira entre os objetos aos quais se aplica e os objetos aos quais não se aplica. A adição de um único fio de cabelo poderia fazer com que uma pessoa careca se tornasse não careca, e a remoção de um único fio poderia fazer com que uma pessoa não careca se tornasse careca. “Careca” seria um predicado preciso. Esta conclusão é absurda, dado que nenhum predicado vago é preciso. Ao contrário, um predicado é vago *somente se* for impreciso, no sentido de não estabelecer qualquer fronteira do tipo acima.

Desta breve reflexão sobre o paradoxo sorites, podemos extrair três critérios de adequação para uma teoria ideal da vagueza. São eles:

- (a) *Critério do sorites*: resolver o paradoxo sorites.
- (b) *Critério da coerência*: não implicar que predicados vagos são incoerentes.
- (c) *Critério da precisão*: não implicar que um predicado pode ser vago e preciso ao mesmo tempo.

Em resumo, é desejável que uma teoria da vagueza resolva o paradoxo sorites sem, com isto, implicar que predicados vagos são incoerentes ou precisos. A falha em satisfazer qualquer um destes critérios deve ser interpretada como um importante defeito de uma teoria. Isto é o máximo que direi sobre estes critérios aqui (uma apresentação e defesa mais cuidadosa pode ser encontrada em Salles (2016)). Como veremos no devido momento, acredito que o Supervalorativismo falha em satisfazer (c). Mas, primeiro, precisamos entender os aspectos básicos desta teoria.

## 2. Supervalorativismo

Como muitas outras teorias da vagueza, o Supervalorativismo afirma que predicados vagos admitem *casos de fronteira*. A ideia de casos de fronteira pode ser melhor entendida em contraste com a de casos claros. Alguns objetos são claramente carecas, são casos claros positivos de

“careca”. Quando *a* é um caso claro positivo de “careca”, a frase “*a* é careca” é verdadeira. Outros objetos são claramente não carecas, são casos claros negativos de “careca”. Quando *a* é um caso claro negativo de “careca”, a frase “*a* é careca” é falsa. Entre os casos claros positivos e negativos, há os casos de fronteiras, aqueles que nem são claramente carecas nem são claramente não carecas. Quando *a* é um caso de fronteira de “careca”, “*a* é careca” é *nem verdadeira nem falsa*. Os casos claros positivos, os casos de fronteira e os casos claros negativos formam, respectivamente, a *extensão positiva*, a *penumbra* e a *extensão negativa* do predicado, de tal modo que seu domínio de aplicação é dividido em três partes.

Até aí, o modelo é similar ao da Teoria Trivalente, discutido em Salles (2019). Entretanto, o que caracteriza um predicado vago não é a tripartição de seu domínio. Supervalorativistas enxergam a existência de casos de fronteira como um sinal de que os predicados vagos são, em algum sentido, incompletos. Eles são incompletos no sentido de que sua extensão não foi completamente determinada, não houve uma decisão sobre qual das precisões será a sua extensão. Apesar de esta imagem já estar pelo menos implícita no clássico artigo de Kit Fine (1975), ela é mais frequentemente expressada pelo que David Lewis chamou de “indecisão semântica”.

The only intelligible account of vagueness locates it in our thought and language. The reason it's vague where the outback begins is not that there's this thing, the outback, with imprecise borders; rather there are many things, with different borders, and nobody has been fool enough to try to enforce a choice of one of them as the official referent of the word 'outback'. Vagueness is semantic indecision (LEWIS, 1986, p. 212).

Aplicada ao caso dos predicados, a tese da indecisão semântica afirma que a vagueza surge do fato de não termos nos decidido por uma dentre as muitas precisões admissíveis de um predicado. O conceito central nesta história é o de *precisão admissível*. Uma *precisão* é a distribuição dos casos de fronteira entre as extensões positiva e negativa do predicado, de modo

que todos os casos de fronteira sejam colocados em uma ou na outra. Embora predicados vagos em geral admitam muitas precisões, nem toda precisão é *admissível*. Certas condições devem ser satisfeitas. Por exemplo, os casos claros devem ser respeitados, nenhuma precisão pode colocar um caso claro positivo na extensão negativa ou um caso claro negativo na extensão positiva. Além disto, precisões devem ser coerentes, nenhuma precisão pode colocar um indivíduo simultaneamente nas extensões positiva e negativa do predicado. De fato, precisões devem respeitar as verdades analíticas, que, muitas vezes, exprimem relações entre diferentes predicados – por exemplo: “Se João é alto, então não é baixo”. Este último ponto implica que o modo como precisamos alguns predicados afeta o modo como podemos precisar outros. Por esta razão, é mais conveniente falar em precisão admissível de uma linguagem como um todo.

A semântica supervalorativista se baseia no conceito de *precisão admissível*, de modo a levar em conta o suposto fato de que há muitas precisões admissíveis de cada predicado vago. O primeiro passo é relativizar as noções clássicas de *verdade e falsidade* às precisões de uma linguagem. Por outras palavras, ao invés de falarmos em verdadeiro/falso *simpliciter*, falamos em verdadeiro/falso *para uma precisão* ou *em uma precisão*. O segundo passo consiste em fornecer as condições de verdade das frases de linguagens vagas em termos de quantificação sobre precisões admissíveis. A partir disto, desenvolve-se noções não relativas de *verdade e falsidade*, usualmente chamadas de “superverdade” e “superfalsidade”, e que alegadamente ocupam o cargo de verdade/falsidade genuína (Keefe, 2000, p.162).

Considere a frase (1) abaixo:

(1) João é careca

Levando-se em conta o fato de que “careca” é vago, há três possibilidades: João é um caso claro positivo de “careca”, João é um caso claro negativo de “careca” ou João é um caso de fronteira. Se João é um caso claro positivo, então toda precisão admissível da linguagem deve

incluí-lo na extensão positiva de “careca”. Neste caso, (1) é verdadeira para toda a precisão admissível da linguagem. Se João é um caso claro negativo, então toda precisão admissível da linguagem deve incluí-lo na extensão negativa de “careca”. Neste caso, (1) é falsa para toda precisão admissível. Por fim, se João é um caso de fronteira, então há precisões nas quais ele é incluído na extensão positiva e precisões nas quais é incluído na extensão negativa do predicado. Neste caso, (1) é verdadeira para algumas precisões e falsa para outras. A ideia dos supervalorativistas foi apresentar as condições de verdade de uma frase vaga como (1) em termos das três possibilidades acima. No primeiro caso (1) é *superverdadeira*, no segundo é *superfalsa* e no terceiro é *indefinida*. Superverdade, superfalsidade e indefinido são entendidos em termos de quantificação sobre precisões admissíveis.

- **Superverdade:** p é superverdadeira se, e somente se, é verdadeira para toda precisão da linguagem.
- **Superfalsidade:** p é superfalsa se, e somente se, é falsa para toda precisão da linguagem.
- **Indefinido:** p é indefinida se, e somente se, é verdadeira para algumas precisões e falsa para algumas precisões<sup>4</sup>.

Esta é a base da semântica supervalorativista para as linguagens vagas. Um dos aspectos curiosos desta semântica é que ela é comparativamente conservadora em relação à lógica clássica. Comece por reparar que na semântica supervalorativista os conectivos não são verofuncionais. Quando ocorre de alguma frase elementar que compõe uma frase complexa ter o valor indefinido, o valor de verdade da frase complexa não é determinado unicamente com base no valor das

---

<sup>4</sup> No caso do Supervalorativismo, o mais recomendado é expor as condições de verdade em termos de frases e não proposições. Isto porque a teoria não deixa claro qual o conteúdo proposicional das frases vagas. Seguramente, podemos dizer que uma frase vaga p expressa uma proposição para cada precisão da linguagem. Mas isto não esclarece qual exatamente é o conteúdo proposicional de p. Pode-se sustentar que o conteúdo de p é fornecido por alguma das proposições particulares que p exprime em alguma precisão, ou que é fornecido por todas elas, ou por nenhuma delas, ou até mesmo que seu conteúdo proposicional é indefinido. Para nossos propósitos aqui, é mais sensato ignorar este problema e falar em termos de frases ao invés de proposições. Uma discussão interessante deste problema pode ser encontrada em Rohrs (2017).

componentes. Suponha que tanto A como B sejam indefinidos. Isto implica que  $\neg A$  também é indefinido. Deste modo, tanto  $(A \vee B)$  como  $(A \vee \neg A)$  são constituídas apenas de disjuntas indefinidas. Apesar disto, a primeira disjunção é ela própria indefinida, enquanto a segunda é superverdadeira. Disto temos que a disjunção não é verofuncional e que uma disjunção pode ser superverdadeira mesmo quando todas as disjuntas são indefinidas. De forma similar, tanto  $(A \wedge B)$  como  $(A \wedge \neg A)$  são constituídas apenas de conjuntas indefinidas. Mas a primeira conjunta é ela própria indefinida, enquanto a segunda é superfalsa. Disto temos que a conjunção não é verofuncional e que uma conjunção pode ser superfalsa mesmo quando todas as conjuntas são indefinidas. Isto é importante porque permite ao supervalorativista reter tanto a lei do terceiro excluído,  $(A \vee \neg A)$ , como a lei da não contradição,  $\neg(A \wedge \neg A)$ . De fato, todo teorema da lógica clássica será mantido (Keefe, 2000, p.163). Isto porque cada precisão da linguagem é um modelo clássico da mesma, no qual os teoremas da lógica clássica são verdadeiros. Cada teorema, portanto, será verdadeiro em toda precisão e, consequentemente, será superverdadeiro. Este é um importante atrativo do Supervalorativismo, principalmente se comparado com outras teorias plurivalentes, como a Teoria Trivalente (TYE, 1994) e o Gradualismo (MACHINA, 1976). Outro atrativo é que a teoria acomoda a tese intuitiva da indecisão semântica. Todavia, é importante ressaltar que a tese da indecisão semântica não implica o Supervalorativismo. De fato, esta tese é compatível com teorias alternativas da vagueza, como o recente e interessante Expressivismo de John MacFarlane (2016). Não é de todo claro que o Supervalorativismo seja o *melhor* modo de acomodar esta tese. Seja como for, não precisamos insistir nisto.

### 3. Supervalorativismo e a solução do sorites

A solução do paradoxo sorites é muito simples, e não nos tomará mais do que um parágrafo. Vimos que uma conjunção pode ser superfalsa mesmo quando nenhuma de suas conjuntas é superfalsa. Existe uma

analogia entre a conjunção e o quantificador universal, de modo que uma frase universalmente quantificada pode ser superfalsa mesmo quando nenhuma de suas instâncias é superfalsa. É precisamente isto que ocorre com a premissa (2):  $\forall n (Ca_n \rightarrow Ca_{n+1})$ . Nenhuma condicional da forma  $Ca_n \rightarrow Ca_{n+1}$  é superfalsa, na medida em que nenhuma terá uma antecedente superverdadeira e conseqüente superfalsa – isto ocorreria somente se  $a_n$  fosse um caso claro positivo e  $a_{n+1}$  um caso claro negativo de “careca”. Deste modo, nenhuma instância de (2) é superfalsa. Apesar disto, (2) é falsa para toda precisão da linguagem. Afinal, para cada precisão da linguagem há um número  $n$  tal que uma pessoa com  $n$  fios de cabelo na cabeça é careca, mas uma pessoa com  $n + 1$  fios não. Ora, o que é falso para toda precisão é superfalso. Portanto, (2) é superfalsa.

#### 4. Supervalorativismo e o critério da precisão

O Supervalorativismo levanta uma objeção bastante direta quanto ao critério da precisão. Assim como há uma analogia entre a conjunção e o quantificador universal, há uma analogia entre a disjunção e o quantificador existencial. Uma disjunção pode ser superverdadeira mesmo quando nenhuma de suas disjuntas é superverdadeira; já uma frase existencialmente quantificada pode ser superverdadeira mesmo quando nenhuma de suas instâncias é superverdadeira. É precisamente isto que ocorre com  $\exists n (Ca_n \wedge \neg Ca_{n+1})$ . Nenhuma conjunção da forma  $Ca_n \wedge \neg Ca_{n+1}$  é superverdadeira, pois nunca ocorre de ambas conjuntas serem superverdadeiras – isto ocorreria somente se  $a_n$  fosse um caso claro positivo e  $a_{n+1}$  um caso claro negativo de “careca”. Em resumo, nenhuma instância de  $\exists n (Ca_n \wedge \neg Ca_{n+1})$  é superverdadeira. Apesar disto,  $\exists n (Ca_n \wedge \neg Ca_{n+1})$  é verdadeira para toda precisão da linguagem e, portanto, superverdadeira. Isto significa que é superverdade que existe um número  $n$  tal que uma pessoa com  $n$  fios de cabelo na cabeça é careca e uma pessoa com  $n + 1$  fios não é careca. Isto parece ser uma violação explícita do critério da precisão.

A principal linha de resposta a esta objeção é a seguinte. O Supervalorativismo nos permite distinguir entre  $(\forall\exists)$  e  $(\exists\forall)$  abaixo. Embora a teoria de fato nos comprometa com  $(\forall\exists)$ , ela nos permite rejeitar  $(\exists\forall)$ . Mas apenas  $(\exists\forall)$  é realmente problemática.

- $(\forall\exists)$  É superverdadeiro que:  $\exists n (Ca_n \wedge \neg Ca_{n+1})$
- $(\exists\forall)$   $\exists n$  tal que é superverdadeiro que  $(Ca_n \wedge \neg Ca_{n+1})$

Repare que  $(\forall\exists)$  afirma algo que todos podemos aceitar: que em toda precisão da linguagem, “careca” terá uma fronteira. Mas  $(\exists\forall)$  afirma algo realmente controverso: que há um número  $n$  tal que, em qualquer precisão da linguagem, uma pessoa com  $n$  fios de cabelo na cabeça é careca e uma pessoa com  $n+1$  fios não. O que realmente importa é rejeitar  $(\exists\forall)$ , e isto o Supervalorativismo nos permite fazer. Não está claro o quão satisfatória é esta resposta. A própria Keefe (2000, p. 183) acredita que o fato de o Supervalorativismo nos comprometer com  $(\forall\exists)$  é um de seus principais defeitos, e Hyde (2009) apresenta objeções interessantes a este tipo de resposta. Em todo caso, não é necessário insistir nisto. Meu objetivo será mostrar que, mesmo que a resposta acima esteja em ordem, o Supervalorativismo não escapa da violação do critério da precisão.

De acordo com o que podemos chamar de “perspectiva tradicional”, um predicado vago divide seu domínio de aplicação em duas (e apenas duas) partes, uma extensão positiva e uma extensão de negativa. Suponha que o domínio de aplicação de “careca” seja o domínio das pessoas. O resultado é que este domínio seria dividido de modo a haver um número  $n$  tal que uma pessoa com  $n$  fios de cabelo é careca e uma pessoa com  $n + 1$  fios não é careca.

Extensão positiva	Extensão negativa
$\{a_0, a_1, a_2, a_3, \dots, a_n\}$	$\{a_{n+1}, a_{n+2}, \dots, a_{10,000}\}$
Verdadeiro	Falso

Ao traçar uma linha divisória entre as extensões positiva e negativa de “careca”, esta perspectiva claramente viola o critério da precisão. O

Supervalorativismo evita esta linha através da postulação de uma penumbra entre as extensões positiva e negativa do predicado. O resultado pode ser representado como segue.

Ext. positiva	Penumbra	Ext. negativa
$ a_0, a_1, \dots, a_n$	$ a_{n+1}, \dots, a_{n+m}$	$ a_{n+m+1}, \dots, a_{10,000} $
Superverdade	Indefinido	Superfalso

Como se nota pela imagem acima, não existe uma fronteira entre as extensões positiva e negativa de “careca”, pois, entre elas, existe uma penumbra contendo os casos de fronteira. Entre o último caso claro positivo e o primeiro caso claro negativo de “careca”, existe um conjunto de casos de fronteira. É verdade, diria o supervalorativista, que existem muitas precisões admissíveis de “careca”, muitos modos de redistribuir os casos de fronteira e traçar uma linha divisória entre as extensões positiva e negativa. Mas isto só confirma o fato de que “careca”, e qualquer outro predicado vago, é impreciso.

Podemos conceder que o supervalorativista evita a existência de uma fronteira entre as extensões positiva e negativa, entre os casos claros positivos e negativos. Mas o preço que ele paga por isto é a postulação de fronteiras que são igualmente inaceitáveis. Na imagem supervalorativista acima, existe um último caso claro positivo de “careca”,  $a_n$ , seguido de um primeiro caso de fronteira,  $a_{n+1}$ ; e existe um último caso de fronteira,  $a_{n+m}$ , seguido de um primeiro caso claro negativo,  $a_{n+m+1}$ . A principal diferença com respeito à imagem tradicional é que agora temos não apenas uma, mas duas fronteiras no domínio de aplicação de “careca”. Ao invés de uma fronteira entre as extensões positiva e negativa de “careca”, agora temos uma fronteira entre a extensão positiva e a penumbra, e outra fronteira entre a penumbra e a extensão negativa. O problema é que, como Sainsbury (1991) corretamente nota, cada uma destas novas fronteiras é tão inaceitável quanto aquela postulada pela perspectiva tradicional. O conceito de *imprecisão* foi introduzido para expressar o sentido em que os

predicados vagos *não possuem* fronteiras. Dado que a existência de uma fronteira entre a extensão positiva/negativa e a penumbra de um predicado vago é tão inaceitável quanto a existência de uma fronteira entre a extensão positiva e a negativa, o conceito de *imprecisão* deve ser incompatível com a existência de ambos os tipos de fronteira. Deste modo, o Supervalorativismo falha em reconhecer a imprecisão dos predicados vagos e, portanto, viola o critério da precisão.

Em certo aspecto, o Supervalorativismo pode ser ainda mais problemático do que a perspectiva tradicional. Refiro-me ao fato de que o *problema fundacional da precisão* aparece com mais força para os supervalorativistas. Este problema pode ser resumido na seguinte pergunta: se predicados vagos são precisos, então como suas fronteiras são determinadas? Se estou certo quanto ao fato de que este problema é um desafio para a perspectiva tradicional (SALLES, 2016, seção 2.8), então ele é um desafio ainda maior para o Supervalorativismo. Afinal, supervalorativistas terão de explicar como cada uma das *duas* fronteiras postuladas por eles é determinada. Por outras palavras, eles terão de explicar que tipo de mecanismo garante que “careca”, por exemplo, trace uma fronteira entre sua extensão positiva e penumbra, e outra entre sua penumbra e extensão negativa; de tal modo que um único fio de cabelo pode determinar *de qual lado* uma pessoa está.

De acordo com Rosanna Keefe (2000), a violação do critério da precisão pode ser evitada se aceitarmos que a própria linguagem na qual o Supervalorativismo é formulada é vaga<sup>5</sup>:

So ‘admissible specification’ is vague and since it is in the proper part of the metalanguage, that metalanguage is vague in the desired way. There could thus be a borderline case admissible specification. Sentences that are false in such a specification but true in all definitely admissible specifications will be borderline cases of ‘true on all complete and admissible specifications’; and “‘p’” is true on all complete and admissible specifications’ will then be indeterminate. The truth-conditions provided by supervaluationism will not

---

<sup>5</sup> Note que Keefe não usa a expressão “critério da precisão”.

then definitely determine any truth-value for  $p$  or definitely determine that the sentence lacks a truth-value. For the indeterminacy over whether the truth-condition is fulfilled implies that we should not conclude that  $p$  is true, but nor should we call  $p$  neither true nor false as we would do if the condition (determinately) failed to be fulfilled. The truth-value status of  $p$  (whether it is true, false or lacks a value) remains unsettled. And sharp boundaries between the true predications and the borderline cases of ‘tall’ are avoided (KEEFE 2000, p. 203).

A proposta é bastante clara. Começamos por assumir que “precisão admissível” é vago. Lembre-se que superverdade, indefinido e superfalsidade são definidos em termos de quantificação sobre precisões admissíveis. Deste modo, “superverdade”, “indefinido” e “superfalsidade” são elas próprias vagas. Estas expressões estão presentes na linguagem de segunda ordem utilizada para explicar a vagueza da linguagem de primeira ordem, e assim a linguagem de segunda ordem também é vaga. Como esperado, sua vagueza é explicada em uma linguagem de terceira ordem que é ela própria vaga, precisamente por empregar as expressões “superverdade”, “indefinido” e “superfalsidade”. O processo continua indefinidamente, formando uma hierarquia infinita de metalinguagens, em que a vagueza de uma linguagem de ordem  $n$  é sempre explicada por uma linguagem de ordem  $n+1$ , que é vaga por conter as mesmas expressões problemáticas (“superverdade” etc.).

Como, afinal, isto evita a violação do critério da precisão? O problema anterior surgiu porque o Supervalorativismo parecia implicar a existência de uma fronteira entre a extensão positiva/negativa e a penumbra de um predicado vago. Dado que a extensão positiva/negativa é formada pelos objetos aos quais a aplicação do predicado gera uma frase superverdadeira/superfalsa, o resultado é a existência de uma fronteira entre os casos aos quais a aplicação do predicado gera uma superverdade/superfalsidade e os casos de fronteira. Tão logo assumimos que “superverdade”, “indefinido”, “superfalsidade” são vagos, as fronteiras acima se dissipam. Não haverá, por exemplo, qualquer transição brusca entre os casos aos quais a aplicação de “careca” gera uma

superverdade e os casos de fronteira, pois haverá uma penumbra entre eles. Podemos ilustrar o ponto com um exemplo. Suponha que  $Ca_n$  é verdadeira para todas as precisões claramente admissíveis da linguagem. Entretanto, há uma precisão  $P$  que nem é claramente admissível nem claramente não admissível, e  $Ca_n$  é falsa em  $P$ . Será  $Ca_n$  superverdadeira? A resposta depende de se “toda precisão admissível” inclui ou não inclui  $P$ . Se incluir, então  $Ca_n$  é indefinida; se não incluir, então  $Ca_n$  é superverdadeira. Uma vez que não houve uma decisão sobre se  $P$  deve ou não ser incluído no escopo de “toda precisão admissível”,  $Ca_n$  é nem claramente superverdadeira nem claramente indefinida; temos aqui um caso de fronteira entre “superverdade” e “indefinido”. Como esperado, o mesmo ocorre com “superfalsidade” e “indefinido”. Ora, se há casos de fronteira entre “superverdade/superfalsidade” e “indefinido” – e se os há em qualquer nível que consideremos da hierarquia – então não há qualquer *corde cirúrgico* entre os casos aos quais a aplicação do predicado gera uma frase superverdadeira/superfalsa e os casos aos quais gera uma frase indefinida. Por outras palavras, não há qualquer linha divisória entre a extensão positiva/negativa e a penumbra do predicado. Predicados vagos são imprecisos.

Ricardo Santos (2010, p. 214) nota que esta estratégia tem um curioso *calcanhar de Aquiles*. O problema é que, dado critérios aceitos pela própria Keefe, sua estratégia implica na inadequação semântica do Supervalorativismo. Ao argumentar contra a Teoria Trivalente da Vagueza, Keefe sugere o seguinte.

The use of a three-valued logic for a vague language requires the assumption that the three values provide an exclusive and exhaustive classification of declarative sentences; if not, it suffers the same defects as the rejected two-valued system. [...] A three-valued logic for a language is inadequate if it is not true that all its sentences take one of the three values (KEEFE, 2000, p. 122).

O ponto que nos interessa é que teorias semânticas devem respeitar o seguinte critério de adequação: o conjunto de valores de verdade apresentado por elas tem de representar uma distinção exaustiva. Se

houver uma frase declarativa na linguagem objeto que não possui qualquer valor do conjunto, então a semântica é inadequada. O conjunto de valores proposto pela semântica supervalorativista é o superverdadeiro, o superfalso e o indefinido:  $\{V, F, I\}$ . Eis o problema: a estratégia da hierarquia infinita de metalinguagens implica que há frases declarativas que não possuem qualquer um destes valores. Isto significa que, pelos próprios critérios de Keefe, o Supervalorativismo seria inadequado.

Retorne ao exemplo de  $Ca_n$  acima. Obviamente,  $Ca_n$  não é superfalsa e, portanto, não é claramente superfalsa. Além disto, vimos que  $Ca_n$  nem é claramente superverdadeira nem é claramente indefinida. Resultado:  $Ca_n$  não é claramente superverdadeira, não é claramente indefinida e não é claramente superfalsa. Ora, este é precisamente o tipo de fato que deveria nos levar a concluir que o conjunto  $\{V, F, I\}$  não é exaustivo. A razão contra a exaustividade do conjunto  $\{V, F\}$  – a razão a favor da inclusão de  $I$  – é que há frases que nem são claramente  $V$  nem são claramente  $F$ . Por paridade argumentativa, o fato de que há frases declarativas da linguagem natural –  $Ca_n$  é uma frase declarativa da linguagem natural – que não são claramente  $V$ ,  $F$  ou  $I$  deveria nos levar à conclusão de que  $\{V, F, I\}$  não é exaustivo. Mas se  $\{V, F, I\}$  não é exaustivo, então a semântica supervalorativista não é adequada. Pelos critérios assumidos pela própria Keefe, o Supervalorativismo é inadequado.

A estratégia da hierarquia infinita de metalinguagens ainda enfrenta outros problemas interessantes. Para que esta estratégia funcione, um predicado vago não pode ser definido meramente como um predicado que admite muitas precisões (no sentido supervalorativista de “precisão”). Suponha que admitir muitas precisões seja suficiente para a vagueza de  $F$ . Neste caso, um predicado  $F$  que admita muitas precisões é vago independentemente de a expressão metalinguística “precisão admissível de  $F$ ” ser ou não vaga. A princípio, nada impede  $F$  de satisfazer as duas condições que seguem: (i)  $F$  é vago, (ii) “precisão admissível de  $F$ ” não é vago. Neste caso,  $F$  faria um corte *cirúrgico* entre sua extensão positiva e

sua penumbra; e outro corte entre sua penumbra e sua extensão negativa. Mas isto significaria que F é preciso. Conclusão: um predicado poderia ser vago e preciso ao mesmo tempo, e o critério da precisão é violado. Para evitar este problema, Keefe deve assumir o seguinte: F é vago somente se (i) F admite muitas precisões e (ii) “precisão admissível de F” admite muitas precisões; “precisão admissível de “precisão admissível de F”” admite muitas precisões; e assim por diante *ad infinitum*. De acordo com esta perspectiva, uma condição necessária para que um predicado ordinário de primeira ordem seja vago é que um predicado metalinguístico de trecentésima vigésima segunda ordem admita muitas precisões. A vagueza da linguagem natural dependerá de fatos sobre uma linguagem artificial de trecentésima vigésima segunda ordem. Isto soa bastante suspeito, para não dizer altamente implausível.

Um último problema é que muitos predicados vagos ordinários simplesmente não admitem o tipo de hierarquia infinita exigida por Keefe. O predicado “careca” é por si próprio vago. Sua vagueza não é meramente derivativa de predicados relacionados, como “fio de cabelo”, “pessoa”, “cabeça” etc. Mesmo que os últimos fossem precisados, “careca” permaneceria vago. De acordo com a estratégia da hierarquia de metalinguagens, isto significa que mesmo que os últimos fossem precisados, a explicação da vagueza de “careca” exigiria uma hierarquia infinita de metalinguagens. Parece-me que podemos mostrar que esta suposição é falsa. Considere a seguinte sequência:  $\langle Ca_0, Ca_1, \dots, Ca_{5,000}, \dots, Ca_{10,000} \rangle$ .  $Ca_0$  é superverdadeira, e  $Ca_{10,000}$  é superfalsa. Suponha que exista pelo menos um item na sequência que seja indefinido. Isto significa que, após a classificação imposta por nossa linguagem supervalorativista de segunda ordem, restará um máximo de 9.998 itens a serem classificados pela linguagem supervalorativista de terceira ordem. Suponha que pelo menos um deles seja um caso de fronteira entre “superverdade” e “indefinido”, e pelo menos um deles seja um caso de fronteira entre “indefinido” e “superfalsidade”. Isto significa que, após a classificação imposta pela linguagem de terceira ordem, ficaremos com um máximo de

9.996 itens não classificados. Na medida em que prosseguirmos com a classificação, as possibilidades em aberto diminuirão. Como o número de objetos a serem classificados é finito, em algum momento todos estarão classificados. Neste momento não poderemos mais ascender a uma ordem superior, teremos chegado a um fim. Por si próprio, o predicado “careca” não admite uma hierarquia infinita de metalinguagens. Keefe teria de lidar com isto dizendo que ou “careca” não é vago, ou é apenas derivativamente vago (vago em virtude da vagueza de outros predicados). O mesmo problema apareceria para outros predicados vagos, como “monte”. Na medida em que tanto “careca” como “monte” são exemplos paradigmáticos de predicados genuinamente (e não apenas derivativamente) vagos, este resultado é implausível.

Em resumo, a estratégia da hierarquia infinita de linguagens não permite ao supervalorativista escapar da violação do critério da precisão. Uma vez que ela é a melhor opção disponível, o Supervalorativismo dificilmente escapa da violação deste critério.

Tenho me deparado com duas perguntas frequentes quando discuto os argumentos acima. A primeira é esta: O supervalorativista não poderia simplesmente rejeitar o critério da precisão? Por “rejeitar o critério da precisão” entendo aqui rejeitar a afirmação de que um predicado é vago somente se é linguisticamente impreciso, no sentido de não possuir uma fronteira entre os casos aos quais se aplica e os casos aos quais não se aplica. Penso que o supervalorativista dificilmente poderia fazer isto. O problema é que, ao fazê-lo, ele tornaria sua teoria menos atraente do que o Epistemicismo. Em suas versões contemporâneas (por exemplo, Williamson (1994)), O Epistemicismo é uma perspectiva atraente em parte por ser mais simples do que suas alternativas plurivalentes. Um exemplo disto é que sua explicação da vagueza é conservadora em relação à lógica clássica, mesmo quando comparada ao Supervalorativismo. Na verdade, o Epistemicismo não implica qualquer revisão em nossa lógica e semântica clássicas. Deste modo, ele alegadamente evita um conjunto de resultados contraintuitivos de suas concorrentes. Seu principal problema é violar o

critério da precisão no sentido acima. Ao seguir o caminho de rejeitar o critério da precisão, um supervalorativista – ou um defensor de qualquer outra teoria plurivalente – abre mão de sua principal vantagem sobre o Epistemicismo, enquanto todas as desvantagens são mantidas. Este é o caminho para uma provável vitória do Epistemicismo.

Passemos então à segunda pergunta: Por que um supervalorativista não pode simplesmente definir um predicado impreciso como um predicado que admite muitas precisões? Se ele fizer isto, poderá protestar que sua teoria de fato satisfaz o critério da precisão, na medida em que nenhum predicado poderá ser simultaneamente vago e preciso. A resposta a esta pergunta já foi antecipada. Uma definição deste tipo seria arbitrária. Isto porque não há mais razão para aceitar as fronteiras então postuladas pelo supervalorativista do que há para aceitar aquelas postuladas pela perspectiva tradicional. Na medida em que temos razão acreditar que a imprecisão dos predicados vagos é incompatível com as fronteiras postuladas pela perspectiva tradicional, também temos razão para acreditar que ela é incompatível com as fronteiras postuladas pelo Supervalorativismo.

Uma saída seria rejeitar que o Supervalorativismo forneça algo como a natureza dos predicados vagos. O Supervalorativismo seria *apenas* uma “técnica modelo-teorética da matemática” (GARCIA-CARPINTERO, 2009, p. 346), e a natureza da vagueza seria fornecida pela teoria filosófica da indecisão semântica. Mesmo que o Supervalorativismo seja desta forma limitado, no entanto, o problema permanece. Como vimos, há outros modelos para a indecisão semântica, como o já citado Expressivismo. Se o Supervalorativismo acomoda a indecisão semântica sem, ao mesmo tempo, acomodar a imprecisão dos predicados vagos, então temos uma boa razão para rejeitar que esta técnica modelo-teorética seja a técnica adequada ao fenômeno da vagueza.

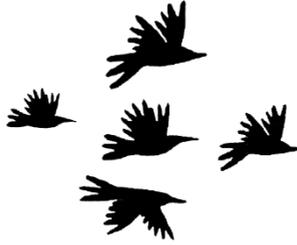
## Conclusão

A principal conclusão deste artigo é que o Supervalorativismo falha em satisfazer o critério da precisão. Apesar de considerar que a falha em satisfazer algum dos três critérios de adequação é um importante defeito de uma teoria da vagueza, não penso que seja o bastante para refutá-la. Isto porque, no fim das contas, não é de todo claro que os três critérios possam ser simultaneamente satisfeitos. Pode-se argumentar que qualquer teoria violará algum deles, e isto deixa em aberto a possibilidade de o Supervalorativismo ainda ser a melhor opção disponível no mercado. Não penso que este seja o caso, em parte porque não acredito que o Supervalorativismo supere o Epistemicismo, e em parte porque penso que é possível construir uma teoria que satisfaz todos os critérios de adequação.

## Referências

- FINE, K. Vagueness, Truth and Logic. **Synthese**, 30, p. 265-300, 1975.
- GARCIA-CARPINTERO, M. Supervaluationism and the report of vague. In: R. Dietz, & S. Moruzzi (Eds.). **Cuts and Clouds**. Oxford: Oxford University Press, 2009, p. 345-59.
- HYDE, D. The prospects of a paraconsistent response to vagueness. In: D. Richard & S. Moruzzi (Eds.). **Cuts and clouds: vagueness, its nature, and its logic**. Oxford: Oxford University Press, 2009, p. 385-405.
- \_\_\_\_\_. The sorites paradox. In: G. Ronzitti. **Vagueness: a guide**. Springer Science & Business Media, 2011, p. 1-17.
- KEEFE, R. **Theories of vagueness**. Cambridge: Cambridge University Press, 2000.
- LEWIS, D. **On the plurality of worlds**. Basil Blackwell, 1986.
- MACHINA, K. F. Truth, belief, and vagueness. **Journal of Philosophical Logic**, 5, p. 47-78, 1976.

- MACFARLANE, J. I-Vagueness as indecision. **Aristotelian Society Supplementary Volume**, 90, 1, p. 255-83, 2016.
- ROHRS, B. Supervaluational propositional content. **Synthese**, 194, 6, p. 2185-201, 2017.
- SAINSBURY, M. Is there higher-order vagueness? **The Philosophical Quarterly**, 41, 163, p. 167-82, 1991.
- SALLES, S. O problema da vagueza. **Fundamento**, n. 12, p. 139-74, 2016.
- \_\_\_\_\_. A Teoria Trivalente da Vagueza e o Problema da Precisão. In: Marcus José Alves de Souza; Maxwell Morais de Lima Filho. (Org.). **Escritos de filosofia III: linguagem e cognição**, Porto Alegre: Editora Fi, 2019, p. 184-200.
- SANTOS, R. O supervalorativismo e a vagueza de ordem superior. In: H. Brito, **Filosofia e literatura 1**. Lisboa: Instituto de Filosofia da Linguagem, 2010, p. 197-217.
- SHAPIRO, S. **Vagueness in context**. Oxford: Oxford University Press, 2006.
- SOAMES, S. Vagueness, partiality and the sorites paradox. In: \_\_\_\_\_. **Understanding Truth**. Oxford: Oxford University Press, 1999, p. 203-27.
- TYE, M. Sorites paradoxes and the semantics of vagueness. **Philosophical perspectives**, 8 (Logic and Language), p. 189-206, 1994.
- WEATHERSON, B. Vagueness as indetermination. In: D. Richard & S. Moruzzi (Eds.). **Cuts and clouds: vagueness, its nature, and its logic**. Oxford: Oxford University Press, 2009, p. 77-90.
- WILLIAMSON, T. **Vagueness**. Routledge, 1994.



## Abstração em Frege

*Alessandro Bandeira Duarte*<sup>1</sup>

### I.

Como é bem conhecido, pelo menos até 1903<sup>2</sup>, ano da publicação do segundo volume de *Grundgesetze der Arithmetik* (GGA), Frege defendeu o logicismo em relação à aritmética, a tese segundo a qual as leis básicas desta ciência poderiam ser provadas a partir de definições lógicas de conceitos aritméticos, axiomas lógicos e regras de inferências que preservam a verdade.

O logicismo de Frege, assim penso, é uma consequência natural do processo de aritmetização da análise que ocorreu no século XIX. A ideia de muitos matemáticos desse período era tornar a análise um campo

---

<sup>1</sup> Possui graduação em Filosofia pela Universidade Federal do Rio de Janeiro, mestrado e doutorado em Filosofia pela Pontifícia Universidade Católica do Rio de Janeiro. Leciona na Universidade Federal Rural do Rio de Janeiro (UFRRJ) e tem experiência na área de Filosofia, com ênfase em Filosofia da Matemática, atuando principalmente nos seguintes temas: Princípios de Abstração Fregeanos, Neo-Logicismo e Neo-Fregeanismo.

<sup>2</sup> De acordo com Dummett (1991, p. 5-6), Frege parece ter percebido que a sua solução ao Paradoxo de Russell não funcionava em 1906. Não podemos inferir daí que este foi o ano em que Frege desistiu do logicismo. De fato, é somente em 1924/25, nos artigos póstumos “Numbers and Arithmetic” (FREGE, 1979, p. 277) e “A New Attempt at a Foundation for Arithmetic” (FREGE, 1979, p. 278), que Frege explicitamente abandona o seu logicismo. É difícil avaliar exatamente quando Frege desistiu do logicismo. Relatos de Carnap, que foi seu aluno até 1914, mostram que Frege, ainda nessa época, sustentava que a aritmética poderia ser derivada por meios lógicos. Por exemplo, no livro *Mein Weg in die Philosophie*, Carnap escreve: “No fim do semestre (de 1913), Frege chamou a atenção de que a nova lógica, a qual ele nos apresentou, poderia servir para construir a matemática inteira” (CARNAP, 1999, p. 8, minha tradução).

matemático autônomo<sup>3</sup>. Até então, os teoremas dessa ciência eram “provados” recorrendo-se às evidências ou às intuições geométricas. Nesse sentido, a análise era dependente ou parasitária da geometria<sup>4</sup>.

Uma das primeiras tentativas feitas para mostrar que a análise não dependia de intuições geométricas (ou de qualquer intuição) foi a de Bernard Bolzano<sup>5</sup>, quando, em 1817, ele provou, analiticamente, baseado apenas na sua definição de continuidade, o Teorema do Valor Intermediário. Podemos citar muitos outros matemáticos engajados nesse processo: Weierstrass<sup>6</sup> e Cantor.

Outra questão importante nesse processo de aritmetização da análise era definir os números reais por meios puramente aritméticos. Além disso, era necessário, a partir dessa definição, provar as propriedades<sup>7</sup> pertencentes a estes números sem recorrer à geometria. Em seu livreto, *Continuity and Irrational Numbers* (1872), Dedekind assume os números racionais e suas propriedades como dados e, a partir daí, ele mostra como construir os números irracionais, cuja união com racionais produz os números reais. Não é nosso objetivo entrar nos detalhes do procedimento de Dedekind aqui. O que queremos enfatizar é que a aritmetização sugerida por Dedekind, nesse artigo, só poderia ser bem-sucedida se os

<sup>3</sup> Veja Demopoulos (1995).

<sup>4</sup> A seguinte passagem de Dedekind (1963, p. 1-2) é reveladora: “As professor in the Polytechnic School in Zürich I found myself for first time obliged to lecture upon the elements of the differential calculus and felt more keenly than ever before the lack of a really scientific foundation for arithmetic. In discussing the notion of the approach of a variable magnitude to a fixed limiting value, and especially in proving the theorem that every magnitude which grows continually, but not beyond all limits, must certainly approach a limiting value, I had recourse to geometric evidences... For myself this feeling of dissatisfaction was so overpowering that I made the fixed resolve to keep meditating on question till I should find a purely arithmetical and perfectly rigorous foundation for the principles of infinitesimal analysis”.

<sup>5</sup> Esse teorema afirma o seguinte: seja  $y=f(x)$  uma função contínua - intuitivamente, uma função contínua é aquela na qual qualquer modificação no *input* (no caso,  $x$ ) leva a uma modificação no nosso *output* (no caso, o valor da função  $y$  em  $x$ ) - sobre um intervalo fechado de números reais  $[a,b]$  e  $N$  um número entre os valores da função  $f(a)$  e  $f(b)$ , então existe um número  $c$  que pertence ao intervalo  $[a,b]$ , tal que o valor da função em  $c$  é  $N$ .

<sup>6</sup> Weierstrass provou, analítica e independentemente, o Teorema do Valor Intermediário que, inicialmente, recebeu o nome Teorema de Weierstrass. Contudo, com a redescoberta da prova de Bolzano, este teorema é conhecido hoje por Teorema de Bolzano-Weierstrass.

<sup>7</sup> A propriedade mais importante é que os números reais formam um conjunto contínuo, sem lacunas. Essa propriedade não é satisfeita pelos números racionais. Estes, como os reais, formam um conjunto denso, ou seja, dados quaisquer dois números racionais  $a$  e  $b$ , existe sempre um número racional  $c$  entre eles, o que implica a existência de infinitos números entre  $a$  e  $b$ . Contudo, o conjunto dos racionais tem lacunas, “buracos”. Por exemplo,  $\sqrt{2}$  não corresponde, como já provado há cerca de 2400 anos atrás, a nenhum número racional.

números racionais pudessem ser definidos e suas propriedades pudessem ser derivadas por meios puramente aritméticos. Em outras palavras, se a definição de número racional e suas propriedades dependessem da intuição geométrica (ou de qualquer outro tipo de intuição), o processo de aritmetização da análise estaria fadado ao fracasso. De alguma forma, a análise dependeria da geometria (ou de algum tipo de intuição) e não seria uma ciência autônoma.

Por sua vez, os números racionais podem ser definidos por meio puramente aritméticos a partir dos números naturais. Há algumas formas de se proceder. Primeiro, poderíamos definir os números inteiros como sendo certos pares ordenados de números naturais (correspondendo à diferença entre eles). Depois, poderíamos definir os racionais como sendo certos pares ordenados de inteiros (correspondendo à fração entre inteiros). Outro procedimento é definir, partindo dos números naturais, diretamente os racionais como sendo certos pares ordenados de números naturais. Os inteiros são, então, definidos como sendo racionais de um certo tipo (LANDAU, 1966). Em ambos os procedimentos, é possível provar as propriedades dos números racionais, dentre as quais podemos mencionar: leis comutativas e associativas da adição e da multiplicação, transitividade da relação *maior que* (que também é definida em termos dos naturais), a densidade do conjunto dos racionais etc.

A discussão informal acima tem o único propósito de evidenciar que a aritmetização da análise depende, em última instância, dos números naturais e suas propriedades. O que é mais importante, a autonomia da análise depende da autonomia da aritmética dos números naturais. Se os números naturais são dados pela intuição (geométrica ou não), então a análise é dependente, de alguma forma, da intuição. O projeto imaginado de aritmetização seria malogrado.

De fato, na *Crítica da Razão Pura* (CPR, 1781/1787), Kant argumenta a favor da tese segundo a qual a matemática pura é sintética *a priori*. Isto significa que, embora esta ciência não dependa de fatos empíricos para provar suas proposições, ela é dependente das intuições puras de tempo e

espaço (veja CPR, B 14-16, A 39/ B 55-6). Portanto, em particular, de acordo com Kant, a aritmética dos números naturais dependeria de algum tipo de intuição. Não é totalmente claro se, para Kant, aritmética depende apenas da intuição temporal, ou se esta ciência também depende da intuição espacial. Em *Prolegomena to Any Future Metaphysics* (PFM, 1783, p. 32), Kant escreve: “Arithmetic accomplishes its concept of number by the successive addition of units in time”.

Em CRP (A 142-3 / B 182), há a seguinte passagem:

A imagem pura de todas as quantidades (*quantorum*) para o sentido externo é o espaço, e a de todos os objetos dos sentidos em geral é o tempo. O *esquema* puro da quantidade (*quantitatis*), porém, como *conceito* do entendimento, é o número, que é uma representação que engloba a adição sucessiva da unidade à unidade (do homegêneo). Portanto, o número não é mais do que a unidade da síntese que eu opero entre o diverso de uma intuição homogênea em geral, pelo fato de eu produzir o próprio tempo na apreensão da intuição (KANT, 1997, A142-3/B 182).

Em outra passagem da CRP (A411/B438), ele escreve:

Para agora dispormos a tábua das ideias segundo a das categorias, tomamos em primeiro lugar os dois *quanta* originários de toda a nossa intuição, o tempo e o espaço. O tempo é em si uma série (*e a condição formal de todas as séries*<sup>8</sup>) pelo que, em relação a um presente dado, podem distinguir-se nele *a priori* os *antecedentia*, como condição (o passado) dos *consequentia* (o futuro).

De acordo com Robert Hanna (2002):

Our pure intuition of the infinite unidirectional successive time-series supplies a fundamental semantic condition for arithmetic\* (HANNA, 2002, p. 336)<sup>9</sup>.

A progressão infinita que os números naturais formam depende, de alguma forma, da série infinita dada pela intuição pura do tempo. Não

---

<sup>8</sup> Meu grifo.

<sup>9</sup> *Arithmetic\**, de acordo com Hanna, inclui a aritmética recursiva primitiva e lógica monádica, que é o sistema estudado, ele assume, por Kant.

queremos aqui discutir minuciosamente a filosofia da matemática de Kant, o que nos levaria muito além do escopo do nosso tema. O que queremos salientar com as observações acima é a necessidade de explicar, contra Kant, como a aritmética dos números naturais não depende da intuição.

Um representante do movimento da aritmetização pode escolher dois tipos de respostas: (1) sustentar que Kant está errado quando afirma que a aritmética dos números naturais depende da intuição. Neste caso, ter-se-ia de dar uma explicação de como obtemos conhecimento das proposições da aritmética sem apelar à intuição; ou (2) tentar-se-ia reduzir a aritmética dos números naturais a algo mais básico que não apele explícita ou implicitamente à intuição e, neste caso, afirmar-se-ia que Kant está errado.

Os dois tipos de resposta acima não são equivalentes. É possível argumentar a favor do caráter não-intuitivo da aritmética dos números naturais, sem reduzir estes números (ontologicamente) às coisas mais básicas que não dependem da intuição.

Durante muito tempo, acreditei que Frege tinha optado pela segunda posição. Ou seja, acreditava que Frege tentou justificar seu logicismo por meio da redução dos conceitos aritméticos a conceitos lógicos. Mas, como tentarei argumentar, as coisas são mais complicadas do que parece. A definição contextual de Frege parece querer justificar que nosso conhecimento dos números é não-intuitivo em termos da primeira posição (não há uma redução, pelo menos ontológica). Mas adiante, discutirei isso melhor.

## II.

Em 1879, Frege publica *Begriffsschrift* (BS). Neste livro, temos a primeira axiomatização do cálculo proposicional, do cálculo de predicados de primeira ordem e do cálculo de predicados de segunda ordem (este implicitamente). Na parte I do livro, Frege introduz e explica seus conceitos lógicos básicos: conteúdo conceitual, negação, implicação, identidade de conteúdo, função, argumento, quantificação universal, além

de introduzir dois símbolos puramente sintático – traço do juízo e o traço de conteúdo. Além disso, a regra de inferência, *modus ponens*, é introduzida e explicada. Implicitamente, Frege assume outras regras de inferência: a regra de substituição uniforme para conteúdos conceituais e para funções, a regra de generalização universal e a regra do confinamento da generalidade ao consequente.

Na parte II de BS, Frege enuncia os axiomas lógicos expressos por meio de seus primitivos lógicos e deriva, usando as regras de inferência, vários teoremas lógicos. Na parte III de BS, Frege define, a partir dos seus primitivos lógicos, conceitos aritméticos de hereditariedade<sup>10</sup>, de ancestral forte<sup>11</sup>, de ancestral fraco<sup>12</sup> e de relação unívoca (função)<sup>13</sup>. E usando os teoremas derivados na parte II, junto com as definições e as regras de inferência, ele prova importantes teoremas sobre a teoria das sequências. Os três mais importantes são o teorema 81 (indução matemática geral), o teorema 98 (transitividade do ancestral forte) e 133 (se  $f$  é funcional, então  $f$  conecta quaisquer dois objetos  $m$  e  $y$  com os quais  $x$  estão na relação de  $f$ -ancestralidade forte).

O que Frege faz em BS é uma redução do conceito seguir-em-uma-sequência ao conceito de implicação (consequência) lógica. Isto significa que Frege, neste livro, está seguindo o tipo de resposta (2) a Kant.

### III.

Em 1884, Frege publica os *Fundamentos da Aritmética*. Nesse livro, Frege busca continuar o trabalho enunciado em BS:

---

<sup>10</sup> Uma propriedade  $F$  é hereditária em uma relação  $f$  justamente quando para quaisquer objetos  $x$  e  $y$ , se  $x$  é  $F$  e  $x$  está na relação  $f$  com  $y$ , então  $y$  é  $F$ .

<sup>11</sup>  $x$  é um  $f$ -ancestral forte de  $y$  justamente quando  $y$  tem toda propriedade  $F$  que  $x$  tem e que é hereditária na relação  $f$ . De acordo Frege, essa relação pode ser traduzida por  $y$  vem depois de  $x$  na relação  $f$ .

<sup>12</sup>  $x$  é um  $f$ -ancestral fraco de  $y$  justamente quando  $x$  é um  $f$ -ancestral forte de  $y$  ou  $x$  é igual a  $y$ . Essa relação pode ser traduzida por:  $y$  pertence à  $f$ -série iniciada por  $x$ .

<sup>13</sup> Uma relação  $f$  é funcional justamente quando para quaisquer objetos  $x$ ,  $y$  e  $z$ , se  $x$  está na relação  $f$  com  $y$  e  $z$  está na relação  $f$  com  $y$ , então  $x=z$ .

Quando me coloquei a questão, a saber, à qual destas duas classes pertencem os juízos da aritmética, tive de testar primeiro o quão longe se poderia chegar na aritmética somente por meio de inferências que dependem das leis do pensamento que transcendem todas as particularidades. Para isto, eu procedi da seguinte maneira: primeiro eu tentei reduzir o conceito de ordenação-em-uma-sequência ao de consequência lógica e disto progredir para o conceito de número (FREGE, 1998, p. X)<sup>14</sup>.

A aritmética, como eu disse no início, foi o ponto de partida da cadeia de pensamentos que me levou a minha notação conceitual. Portanto, eu pretendo aplicá-la, primeiramente, a esta ciência, tentando analisar seus conceitos em seguida e estabelecer mais profundamente seus teoremas. Por enquanto, eu apresentei na terceira parte algumas coisas que se movem nesta direção. Outros prosseguimentos do curso sugerido – a elucidação do conceito de número, de magnitude e assim por diante – devem constituir o assunto de outras investigações que pretenderei produzir imediatamente após este livro (FREGE, 1998, p. XIV)<sup>16</sup>.

Nos *Fundamentos da Aritmética*, Frege propõe então elucidar o conceito de número cardinal (*Anzahl*). Podemos dividir este livro em duas partes: a primeira negativa, Frege discute e refuta uma série de opiniões referentes à natureza das proposições aritméticas, ao conceito de número, à noção de cardinalidade. Embora a primeira parte seja negativa em caráter, algumas teses Fregeanas estão implicitamente localizadas nestas seções: por exemplo, nas §§ 5-8 e §18, Frege argumenta a favor da derivabilidade das proposições aritméticas, da possibilidade de se definir (e obter) os números naturais individuais seguindo a proposta de Leibniz. Nas §§9-17, Frege rejeita que as leis da aritmética sejam verdades indutivas (contra Mill), e ele apresenta argumentos a favor da analiticidade (e do caráter lógico) do conceito de número (e da aritmética). Incrivelmente, o argumento de Frege não é que números são extensões de conceitos que são objetos lógicos por natureza, mas:

---

<sup>14</sup> FREGE (1972, p. 104).

<sup>15</sup> Todas as passagens de Frege foram traduzidas por mim.

<sup>16</sup> FREGE (1972, p. 107).

O fundamento da aritmética não é mais profundo que o de todo saber empírico, mais profundo mesmo que o da geometria? As verdades aritméticas governam o domínio do enumerável. Este é o mais inclusivo; pois não lhe pertence apenas o efetivamente real, não apenas o intuível, mas todo o pensável. Não deveriam, portanto as leis dos números manter com as do pensamento a mais íntima das conexões? (FREGE, 1988, p. 29)<sup>17</sup>.

Nas §§18-39, Frege discute várias concepções de números defendidas por seus contemporâneos ou por autores clássicos. O que Frege apresenta de positivo destas discussões está sumarizado na §45: (1) números não são abstraídos de coisas na maneira que cor, peso são, número não é uma propriedade de objeto; (2) número não é algo físico, caso contrário, em que sentido ele teria uma aplicação universal?; (3) número não é a resultante de adicionar coisa a coisa; (4) os termos, conjunto, pluralidade, multiplicidade são inadequados (e vagos) para usar na definição de número; (5) o conceito de número parece encerrar em si duas propriedades contraditórias: o um e o muitos.

Como resposta às questões levantadas nestas seções, Frege sugere que “uma atribuição numérica contém uma predicação de um conceito” (FREGE, 1988, p. 60)<sup>18</sup>. Isto explicaria porque números diferentes podem ser atribuídos a uma mesma coleção de coisas, da mesma maneira explicaria a aplicabilidade universal do número, uma vez que sob conceitos podem cair o que é físico e mental, o que temporal e espacial e o que é não-temporal e não-espacial. Também explica como a contradição mencionada em (5) acima não é, de fato, uma contradição. Conceitos tem um critério de aplicação e este critério informa a unidade em questão (Aristóteles). Por outro lado, alguns conceitos (sortais) têm um critério de identidade que se aplica aos objetos que caem sob eles. Daí a diversidade. Não há contradição.

---

<sup>17</sup> FREGE (1980, p. 21).

<sup>18</sup> FREGE, 1980, p. 59.

#### IV.

Acredito que, inicialmente, Frege tentaria, nos *Fundamentos da Aritmética*, dar uma resposta a Kant em termos de (1) acima. Mas, quando ele estava escrevendo o livro, ele percebeu que algo não estava certo e ele tentou corrigir. O resultado é a tensão mencionada. Quero apresentar alguns fatos para apoiar minha interpretação. Em 1882, Frege escrevera uma carta a Anton Marty (ou Carl Stumpf) na qual ele anuncia o seguinte:

Agora, eu estou quase completando um livro no qual eu trato do conceito de número cardinal [Anzahl] e demonstro que os primeiros princípios sobre contar os números [*ersten Sätze über das Zählen der Zahl*], que até agora foram considerados, em geral, como axiomas indemonstráveis, podem ser provados a partir de definições por meio de leis lógicas somente, de maneira que estes princípios podem ser considerados como juízos analíticos no sentido de Kant (FREGE, 1976, p. 163).

Este suposto livro foi escrito na notação conceitual, como a seguinte passagem sugere:

Eu me encontro em um círculo vicioso: antes que as pessoas deem atenção à minha notação conceitual, elas querem ver o que eu posso fazer com ela e eu, por minha vez, não posso mostrar isto sem pressupor uma familiaridade com a minha notação conceitual. Assim, parece que dificilmente contarei com qualquer leitor para o livro que mencionei no início [da carta] (FREGE, 1976, p. 165).

Alguns dias depois de enviar esta carta, Frege recebera uma correspondência de Carl Stumpf na qual este último sugere que Frege escreva um livro na linguagem ordinária e, somente depois, publique o livro na notação conceitual. Frege parece ter seguido a sugestão de Stumpf e nasceu daí os *Fundamentos da Aritmética*.

Infelizmente, não existe nenhuma cópia desse manuscrito de Frege. Talvez tenha se perdido junto com os demais na segunda guerra mundial. Mas, parece ser razoável supor que os *Fundamentos da Aritmética* tivesse

algum tipo de ligação com este manuscrito. Aqui surge a nossa perplexidade.

## V.

Na introdução dos *Fundamentos*, Frege anuncia pela primeira vez o Princípio do Contexto, a tese segunda a qual nunca se deve perguntar pelo significado (*Bedeutung*) de uma palavra isoladamente, mas apenas no contexto de uma sentença. Além disso, Frege anuncia outros dois princípios: separar nitidamente o psicológico do lógico, o subjetivo do objetivo e nunca perder de vista a distinção entre conceito e objeto. Como anunciado na introdução, estes três princípios são metodológicos e interdependentes. Quando tomamos o significado dos numerais isoladamente, poderia ser o caso de ligarmos a este significado uma ideia. Assim, violaríamos o primeiro princípio. Por outro lado, no contexto de uma sentença, saberíamos o *status* ontológico (significado) de uma expressão. Por exemplo, se uma expressão ocorre como o sujeito de uma sentença que expressa um conteúdo singular (sentença atômica), então tal expressão significará um objeto. Se outra expressão for um predicado em tal sentença, então esta expressão significará um conceito.

A segunda ocorrência do Princípio do Contexto é na §60: “é suficiente que uma proposição como um todo tenha um sentido para que seja conferido um conteúdo às suas partes” (FREGE, 1988, p. 70)<sup>19</sup>. Aqui, parece, temos uma formulação ontológica do Princípio do Contexto, uma vez que é uma condição para existência de conteúdo das partes da sentença.

A terceira ocorrência do Princípio do Contexto é na §62 (terei mais a dizer). A quarta ocorrência do princípio encontra-se na §106, que está intimamente ligada com a §62.

---

<sup>19</sup> FREGE (1980, p. 71).

## VI.

Na §62, depois de rejeitar sua primeira tentativa de definição do conceito de número (uma vez que ela não dava conta do caráter “ontológico” de números como objetos), Frege evoca o Princípio do Contexto na sua segunda tentativa de definição:

Como nos pode então ser dado um número, se não podemos ter dele nenhuma representação ou intuição? *Apenas no contexto de uma proposição as palavras significam algo. Importará, portanto, definir o sentido de uma proposição onde ocorra um numeral.* Por enquanto, isto fica ainda muito a critério de nosso arbítrio. Mas já estabelecemos que se deva entender pelos numerais objetos independentes. Com isto temos uma espécie de proposições que devem ter um sentido, proposições que exprimem um reconhecimento. Se um sinal *a* deve designar para nós um objeto, devemos dispor de um critério para decidir, em qualquer caso, se *b* é o mesmo que *a*, ainda que nem sempre sejamos capazes de aplicá-lo. Em nosso caso, devemos definir o sentido da proposição: O número que convém ao conceito *F* é o mesmo que convém ao conceito *G*; isto é, devemos reproduzir de outra maneira o conteúdo desta proposição, sem empregar a expressão o número que convém ao conceito *F* (FREGE, 1988, p. 71)<sup>20</sup>.

Citando Hume (§63), Frege propõe definir esta identidade em termos de correlação 1-1 entre os objetos que caem sob *F* e *G*. A definição de correspondência 1-1 pode ser expressa em termos puramente lógicos (lógica de segunda ordem). E esta definição não pressupõe a noção de número. Frege adianta uma possível crítica à definição proposta, a saber, de que ele esteja definindo a identidade para o caso especial dos números. Ele afirma que o que está sendo feito é usar “o conceito de identidade, tomado como já conhecido, para chegar naquilo que tem de ser considerado como idêntico” (FREGE, 1988, p. 72)<sup>21</sup>, que no caso são os números. Há uma interessante passagem na seção 63:

---

<sup>20</sup> FREGE, 1980, p. 73.

<sup>21</sup> FREGE, 1980, p. 74.

Esta parece ser decerto uma espécie muito incomum de definição, a qual os lógicos ainda não prestaram suficiente atenção; que ela não é inaudita, alguns exemplos podem mostrá-lo (FREGE, 1988, p. 72)<sup>22</sup>.

O primeiro exemplo de Frege é a transformação de um juízo que expressa a relação de paralelismo entre retas em um juízo que expressa a identidade entre direções de retas. A relação de paralelismo é uma relação de equivalência e, como tal, esta relação expressa algo que é comum às retas quando elas são paralelas. E esta coisa em comum é sua direção. Mas, uma vez que ‘a direção da reta *a*’ é um nome próprio, podemos transformar a relação de paralelismo entre retas em uma identidade de suas direções (que é o tipo próprio de sentença na qual expressões podem significar objetos).

Frege sustenta uma prioridade epistêmica do juízo que expressa paralelismo entre retas em relação ao juízo que expressa a identidade das suas direções. De acordo com ele, direções são objetos não intuitivos, dos quais podemos ter algum conhecimento via relações intuitivas dadas pelas retas. Ele retruca autoridades que defendem que o caminho da definição é o inverso, isto é, definindo que retas são paralelas, justamente quando elas têm a mesma direção. Para Frege, isso inverte “verdadeira ordem das coisas”<sup>23</sup>.

O ponto obscuro é se Frege admitiria uma redução ontológica das direções de retas a retas paralelas? Parece razoável dizer que não, caso contrário, a expressão “a direção da reta *a*” não significaria nada, uma vez que a identidade de direções reduzir-se-ia ao paralelismo entre retas.

Na §65, Frege tenta a seguinte definição:

#### **Princípio de Direção**

$$\forall a \forall b [Dir(a) = Dir(b). \equiv a \parallel b]$$

---

<sup>22</sup> FREGE, 1980, p. 74.

<sup>23</sup> Há uma conexão dessa discussão com a tese de doutorado de Frege sobre geometria projetiva. Veja Wilson (1992, 1999).

Em termos epistemológicos, a proposição ‘reta  $a$  é paralela à reta  $b$ ’ tem prioridade sobre a proposição ‘a direção da reta  $a$  = a direção da reta  $b$ ’, uma vez que temos intuições de retas e da relação de paralelismo. Mas, ontologicamente, assim estou supondo, nenhuma proposição tem prioridade sobre a outra. Não podemos reduzir a nossa fala de identidade de direções de retas à fala de paralelismo entre retas.

Agora, vamos transportar isso para o caso dos números. De acordo com o que foi mencionado acima, a sentença “Existem tantos  $F$ s quantos  $G$ s” (equinumerosidade, correspondência 1-1) tem o mesmo significado (conteúdo conceitual) que “o número do conceito  $F$  = o número do conceito  $G$ ”.

#### **Princípio de Hume**

$$\forall F \forall G [N_x: Fx = N_x: Gx. \equiv. F1 - 1G]$$

Números são objetos não-intuitivos (argumentado anteriormente nos *Fundamentos*). A relação de equinumerosidade (correspondência 1-1) é uma relação lógica (reduzível aos conceitos lógicos propostos em BS). Sendo lógica, esta relação é *a priori* e analítica e independente da intuição. Esta relação tem uma prioridade epistemológica sobre a identidade numérica, contudo não temos uma redução ontológica, caso contrário, não teríamos números como objetos. No Princípio de Hume, números seriam objetos *sui generis*, cuja explicação epistemológica de como obtemos conhecimento sobre eles é dada via nosso conhecimento lógico de correspondência 1-1. Com o Princípio de Hume, eu creio que é o primeiro tipo de resposta mencionada acima que Frege desejava dar a Kant.

Outro ponto importante a ser salientado é que dada a relação de mesmidade de conteúdo (ou significado) entre as sentenças que ocorrem no Princípio de Hume, então se um dos lados da equivalência tiver conteúdo, o outro necessariamente terá e vice-versa (se um lado é verdadeiro, o outro será). Isso, juntamente com a versão ontológica do Princípio do Contexto na §60, garante a existência dos conteúdos das expressões “a direção da reta  $a$ ”, “o número do conceito  $F$ ” (quando as

sentenças que ocorrem à direita dos princípios supracitados tiverem sentido).

## VII.

Todavia, Frege descarta esta tentativa de se definir número também. Ele elabora uma objeção a este tipo de definição que, ele defende, é fatal. Esta objeção é conhecida como o Problema de Júlio César:

Mas, há ainda uma terceira dúvida que pode nos fazer suspeitar da definição proposta. Na proposição ‘a direção de  $a$  é idêntica à direção de  $b$ ’, a direção de  $a$  desempenha o papel de um objeto, e nossa definição nos permite um meio de reconhecer este objeto como o mesmo novamente, no caso deste objeto aparecer repentinamente de alguma outra forma, por exemplo, como a direção de  $b$ . Mas, este meio não é suficiente para todos os casos. Por exemplo, ele não decidirá para nós se a Inglaterra é a mesma que a direção do eixo da Terra — se eu posso ser desculpado pelo exemplo que parece ser sem sentido. Naturalmente, ninguém confundirá a Inglaterra com a direção do eixo da Terra, mas isto não ocorre graças à nossa definição de direção (FREGE, 1988, p. 75)<sup>24</sup>.

Na forma como o problema é apresentado acima, não fica muito claro, pelo menos para mim, o que Frege pretende. Ele não articula totalmente seus argumentos. A questão é que, pela definição de direção de uma reta  $a$ , temos:

A direção de uma reta  $a =$  a direção da reta  $b$  justamente quando as retas  $a$  e  $b$  são paralelas.

Frege assume o conceito de identidade como dado. Uma vez que ‘a direção da reta  $a$ ’ desempenha o papel de um objeto, então podemos, usando o conceito de direção, perguntar se esta direção é igual à direção de outra reta, digamos,  $b$ . Essa sentença tem o mesmo conteúdo que a sentença “ $a$  e  $b$  são paralelas”. Se for o caso dessas retas serem paralelas,

---

<sup>24</sup> FREGE, 1980, p. 77-8.

então suas direções são idênticas. Se não for o caso, suas direções serão diferentes. Mas poderíamos perguntar, como na passagem citada, se a Inglaterra é idêntica a direção da reta  $a$ . Essa é a questão: se admitirmos que o Princípio da Direção é capaz de responder isso, então teríamos que a sentença acima tem o mesmo conteúdo que “a reta  $a$  é paralela à Inglaterra”. Mas o que isso significa?

A coisa é ainda mais aguda com o Princípio de Hume. Sentenças da forma:

o número de  $F$  = o número de  $G$

seriam respondidas afirmativa ou negativamente, conforme se existem tantos  $F$ s quantos  $G$ s ou não. Agora, pergunte se o número de  $F$  = Júlio César. Se assumirmos que o Princípio de Hume pode resolver esta identidade, então temos que a sentença acima tem o mesmo conteúdo que “Existem tantos  $F$  quantos Júlio César”. Mas esta sentença não tem sentido, uma vez que engloba quantificação sobre Júlio César, mas quantificação é um conceito de segunda ordem sob o qual caem conceitos de primeira ordem e Júlio César é um objeto. Assumindo que Princípio de Hume pode dar respostas à identidade acima, chegamos à conclusão que um dos lados não tem sentido, mas então o outro (a identidade) também não terá sentido e assim não lhe será conferido existência a suas partes (*Fundamentos*, §60).

Todavia, se os princípios acima não resolvem estas questões da identidade, qual será o conteúdo delas? Essas sentenças de identidade têm conteúdo? Frege sugere que elas têm conteúdo (“ninguém irá confundir a direção do eixo da Terra com a Inglaterra”). Portanto, a questão inicial: “como obtemos este conteúdo, principalmente quando um dos objetos é não-espacial e não-temporal?” (*Fundamentos*, §62) seria respondida por meio da definição contextual e o princípio do contexto estabelecidos para este propósito. Mas, uma vez que a definição falha, o que temos?

Há outro ponto a ser mencionado. Obviamente, estas definições não são definições no sentido estrito (definições explícitas): elas não satisfazem

nem o critério de eliminabilidade, nem o de não-criatividade, nem deveriam, pois, neste caso, teríamos a redução das expressões que ocorrem no lado esquerdo às que ocorrem no lado direito. Mas, uma vez que estes princípios são criativos, os dois lados dos princípios não podem expressar o mesmo conteúdo. Aqui, estou assumindo que o que Frege entende por conteúdo é a mesma coisa que ele entende por conteúdo conceitual em BS. Frege não tem muito a dizer sobre conteúdos, mas ele afirma quando dois conteúdos conceituais  $C$  e  $C'$  são idênticos: quando, de um mesmo conjunto de sentenças  $T$ ,  $C$  e  $C'$  derivam as mesmas consequências (em uma mesma linguagem). Mas, no caso do Princípio de Hume (o mesmo vale para o Princípio de Direção), o conjunto de consequências de cada lado desse Princípio não é igual (dentro da lógica de segunda ordem). Do lado esquerdo é possível provar a existência de números, fato que não pode ser provado das instâncias do lado direito.

## VIII.

Na §68, Frege introduz, então, as extensões de conceitos. Um fato notável é que Frege não tem muito a dizer sobre extensões de conceitos (apenas em duas passagens). Ele assume que já é conhecido o que é uma extensão de conceito. Além disso, ele define o operador-cardinalidade em termos de extensão: o número que pertence ao conceito  $F$  é a extensão do conceito 'ser equinúmero ao conceito  $F$ '. Uma vez que já sabemos o que é extensão, então já sabemos que Júlio César não é uma extensão. Além disso, Frege já está assumindo que existem as extensões que são objetos. A questão é: qual papel o princípio do contexto desempenha nos *Fundamentos*? Dummett (1991, p. 201-2) sugere um papel metodológico: o princípio do contexto funciona como um guia para formular uma definição correta de número cardinal. Mas, como afirmamos anteriormente, o princípio do contexto parece ter uma interpretação ontológica (§60) e uma epistemológica (§62). Mas se a definição já nos apresenta o objeto (fora do contexto de uma sentença), o que fazemos com

estas interpretações do princípio? Por outro lado, Sluga (1980, p. 73) afirma que o papel do princípio do contexto é legitimar definições contextuais. Mas Frege as rejeita. Legitimar o que, se não há nada para legitimar. Estas são questões difíceis.

## IX.

A minha posição é a seguinte: Frege inicialmente introduziu os números contextualmente via Princípio de Hume (essa não era uma prática totalmente nova, de acordo com Mark Wilson (1991)) no suposto livro mencionada na carta Marty. Ao aceitar a sugestão de Stumpf, Frege, em algum momento, deparou-se com a dificuldade lógica em relação a esta definição. Ele então resolve introduzir as extensões e assumi-las como já conhecidas, solucionando o problema. Isso explica a falta de comprometimento de Frege em relação às extensões (§107):

Nesta definição, o sentido da expressão ‘extensão de um conceito’ é assumido ser conhecido. Esta maneira de superar a dificuldade não pode ser esperada encontrar uma aprovação universal e muitos prefeririam outros métodos de remover a dúvida em questão. Não atribuo qualquer importância decisiva na introdução das extensões de conceitos (FREGE, 1988, p. 106)<sup>25</sup>.

Outro ponto interessante é a estrutura dos esboços de prova das leis aritméticas (axiomas de Peano). Frege esboça a prova do Princípio de Hume (*Fundamentos*, §73) e, a partir daí, este Princípio é o ator principal das demais provas (esboço). Tirando as seções 66-68 e algumas outras poucas passagens, o livro continuaria em ordem. Vale mencionar que, no esboço da prova do Princípio de Hume, Frege implicitamente assume algo como sua lei básica V. A questão é: ele já tinha ideia da necessidade desta lei quando publicou os *Fundamentos*? Se ele tinha, por que então ele expressa a falta de comprometimento com as extensões? Por que ele demorou tanto tempo para publicar as *Leis Básicas da Aritmética* (1893)

---

<sup>25</sup> FREGE, 1980, p. 117.

já que ele tinha um manuscrito completo? Esse manuscrito é o manuscrito que ele afirma ter descartado na introdução das *Leis Básicas*? Certamente, com a introdução dos valores de verdade, a distinção entre sentido e referência e a introdução dos percursos de valores, Frege teria de começar quase do zero (o manuscrito foi escrito em 1882)<sup>26</sup>.

## Referências

CARNAP, R. **Mein Weg in die Philosophie**. Stuttgart: Reclam, 1999.

DEDEKIND, R. **Essays on theory of numbers**. New York: Dover Publications, 1963.

DEMOPOULOS, W. Frege and rigorization of analysis. In: DEMOPOULOS, W. (Ed.). **Frege's philosophy of mathematics**. Cambridge, Massachusetts: Harvard University Press, 1995. p. 29-40.

DUARTE, A. **Lógica e aritmética na filosofia da matemática de Gottlob Frege**. Tese de Doutorado, Rio de Janeiro: PUC-Rio, Junho 2009.

DUMMETT, M. *Frege: Philosophy of mathematics*. Cambridge, Massachusetts: Harvard University Press, 1991.

FREGE, G. **Conceptual notation and related articles**. Traduzido e editado por T. W. Bynum. Oxford: Clarendon Press, 1972.

FREGE, G. **Wissenschaftlicher Briefwechsel**. Editado por Gottfried Gabriel, Hans Hermes, Friedrich Kambartel, Christian Thiel, Albert Veraart, Hamburgo: Felix Meiner Verlag, 1976.

FREGE, G. **Posthumous writings**. Traduzido por Peter Long e Roger White. Oxford: Basil Blackwell, 1979.

FREGE, G. **The foundations of Arithmetic: a logico-mathematical enquiry into the concept of number**. 2a edição revisada. Traduzido por J. L. Austin. Oxford: Basil Blackwell, 1980.

---

<sup>26</sup> De fato, em DUARTE (2009), apresento um quebra-cabeça relacionado à prova do Princípio de Hume. Lá, mostro que Frege não tinha a sua disposição uma fórmula necessária para derivação desse princípio.

FREGE, G. **Die Grundlagen der Arithmetik**: eine logisch Untersuchung über den Begriff der Zahl. Editado por Christian Thiel. Hamburg: Felix Meiner Verlag, 1988.

FREGE, G. **Begriffsschrift und andere Aufsätze**. Editado por Ignacio Angelelli. Hildesheim: Georg Olms Verlag, 1998.

HANNA, R. Mathematics for humans: Kant's philosophy of arithmetic revisited. **European Journal of Philosophy**, v. 10, n. 3, p. 328–352, 2002.

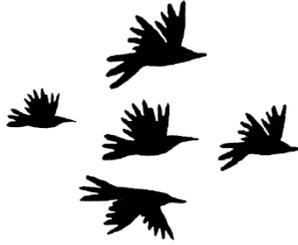
KANT, I. **Crítica da razão pura**. Traduzido por Manuela Pinto Santos e Alexandre Fradique Morujão. Lisboa: Fundação Calouste Gulbenkian, 1997.

LANDAU, E. **Foundations of analysis**. 3.ed. New York: Chelsea, 1966.

SLUGA, H. **Gottlob Frege**. London: Routledge and Kegan Paul, 1980.

WILSON, M. Frege: The Royal road from geometry. **Noûs**, v. 26, n. 2, p. 149–180, 1992.

WILSON, M. To Err is Humean. **Philosophia Mathematica**, v. 7, p. 247–57, 1999.



## Neodescritivismo sobre termos gerais

*Claudio Costa*<sup>1</sup>

Quero aqui esboçar uma teoria neodescritivista dos termos gerais. Minha hipótese é a de que as teorias descritivistas tradicionais dos termos gerais falharam por serem demasiado rudimentares<sup>2</sup>; foi, em parte, devido ao caráter simplista daquelas teorias que a teoria causal da referência dos termos gerais e o externalismo semântico pareceram opções plausíveis, sugerindo que as palavras precisam significar mais do que são convencionadas a significar, posto que, em sua dimensão mais relevante, que é a da determinação da referência, os significados seriam externos a nós mesmos... Contudo, embora ninguém queira negar a indispensabilidade da influência de fatores causais externos para que a referência se torne possível, permanece a questão de se saber se não existe uma falácia genética em se reduzir o significado a algo que (no dizer de Putnam) “está fora da cabeça”<sup>3</sup>. O papel dos fatores causais na realização

---

<sup>1</sup> Professor Titular de Filosofia da Universidade Federal do Rio Grande do Norte (UFRN), com Doutorado em Filosofia pela Universität Konstanz. Realizou Estágios Pós-Doutorais nas seguintes instituições: Hochschule für Philosophie, University of California-Berkeley, Oxford University, Konstanz Universität, Gothemburg Universitet e École Normale Supérieure. Em 2018, publicou o livro “Philosophical Semantics: Reintegrating Theoretical Philosophy” (Cambridge Scholars Publishing), o qual pode ser entendido como uma tentativa de reintegração da filosofia da análise linguística inspirada na releitura da análise do enunciado predicativo singular feita por Ernst Tugendhat.

<sup>2</sup> *E.g.* Hempel (1967).

<sup>3</sup> Hilary Putnam (1975).

da referência dos termos gerais é ainda mais elusivo do que no caso dos nomes próprios.

Um problema é que o terreno que se encontra à frente parece muito mais acidentado do que o da classificação dos termos singulares em indexicais, descrições definidas e nomes próprios. Não há uma única classificação coerente para os termos gerais e parece claro que se há um modelo a comandar as aplicações das regras conceituais, ele deve ser múltiplo. Por isso e porque a questão demanda uma investigação muito mais detalhada, não pretendo fazer mais do que esboçar algumas breves sugestões.

### **Uma classificação para os termos gerais**

Quero começar propondo uma classificação genérica dos termos gerais análoga à dos termos singulares, que como é sabido se dividem em indexicais, descrições definidas e nomes próprios. Correspondentemente, os termos gerais dividir-se-iam no que chamarei de termos *indexicadores*, *descritivadores* e *nominadores*.

Termos gerais indexicadores são os que só se deixam definir em contextos indexicais. Exemplo seriam palavras como ‘vermelho’, ‘redondo’, ‘quente’. Característico desses termos é que eles não se deixam analisar na forma de descrições, ao menos enquanto forem entendidos como designações de estados, aparências fenomenais ou seus correlatos externos. Em contextos perceptuais, esses termos podem ser usados junto a indexicais de modo a determinar a referência sem a intermediação de descrições. Há boas razões para pensar que não há como analisarmos fenomenalmente os termos gerais em proferimentos como “Vermelho ali”, “Redondo lá” ou “Sinto calor” sem aprendermos o significado fenomenal de palavras como ‘vermelho’, ‘redondo’ ou ‘quente’ por alguma coisa como *familiaridade* (*acquaintance*) sensorio-perceptual, isto é, sem sermos apresentados a coisas vermelhas ou redondas ou quentes, sem recorrermos ao uso dessas palavras em proferimentos indexicais. É desse

modo que termos gerais indexadores ganham conteúdo indexical, que a eles se torne possível associar uma regra de aplicação conceitual Interpessoalmente convencionável. Diversamente dos termos singulares indexicais, porém, eles não precisam ser reconhecidos como denotando particulares, a não ser indiretamente, posto que geralmente vem associados a termos ou expressões capazes de identificar particulares. Por exemplo: não posso saber que a propriedade externa do vermelho que vi ontem é a mesma que o vermelho que vejo agora, a menos que a identifique como o vermelho *desse sofá*, associando assim a palavra ‘vermelho’ a uma propriedade particularizada (o *tropo*) do objeto referido pela expressão referencial ‘esse sofá’. Os termos indexadores são os subrogados daquilo que Bertrand Russell equivocadamente chamou de nomes próprios lógicos: afinal, eles são os melhores candidatos a termos primitivos da linguagem, entendidos como aqueles cujas regras de significação só são aprendidas pela familiaridade com os seus *designata*.

Vejamos agora os termos gerais descritivadores. Eles são análogos às descrições definidas, mas com função classificadora ao invés de individuadora. Esses termos aparecem na forma de predicados complexos e podem, sem muita dificuldade, ser traduzidos como descrições indefinidas, como é o caso da descrição ‘...um caçador de dotes’, que pode ser aplicada tanto a Eduardo quanto a Gustavo. Um imenso número de predicções (na verdade, a maioria delas) se caracteriza de forma descritiva, sendo elas mais estáveis que os termos indexadores, na medida em que independem da familiaridade.

Há, finalmente, termos gerais nominadores: *nomes gerais* cuja expressão simbólica perceptível é não-descritiva, mas que são analisáveis de modo em alguma medida análogo ao dos nomes próprios, dado que abreviam descrições ou conjuntos de descrições. Esses são termos de espécies naturais como ‘tigre’, termos de massa como ‘água’, termos de artefatos como ‘cadeira’, termos sociais como ‘professor’, termos culturais como ‘arte’ e ainda muitos termos utilizados em filosofia e ciência. Eles são abreviações de descrições ou de feixes ou de grupos de feixes de descrições

que podem ser mais ou menos estruturados. Ao menos em alguns casos, como veremos, eles podem ser entendidos como abreviações de feixes de regras-descrições cujas combinações efetivamente aplicáveis são selecionadas por meio de regras de atribuição de nível superior, o que não existia no descritivismo tradicional.

Diante desse curioso paralelismo, é possível propor para os termos gerais uma hipótese genético-estrutural. Os termos gerais indexadores seriam estruturalmente mais originários, podendo a cognição das propriedades por eles designadas ser tomada como elemento na construção de conceitos mais complexos. Admitindo uma ontologia dos tropos que, por princípio, rejeita entidades verdadeiramente abstratas, os termos gerais nominadores deveriam, ao menos tendencialmente, ter sua origem em termos gerais descritivadores, que se originariam de combinações de termos indexadores.

Essa analogia classificatória entre termos singulares e gerais nos leva a perguntar se não existiriam situações originárias nas quais eles não se tivessem ainda diferenciado. Podemos imaginar a existência de termos referenciais originários que funcionariam indiferenciadamente, tanto como termos singulares indexicais quanto como termos gerais, em um papel que recorda aquilo que Ernst Tugendhat chamou de *quasi-predicado*.<sup>4</sup> Tratar-se-iam de predicados indexicalmente usados, como o grito de alerta de certas aves, que designariam propriedades singularizadoras pertencentes a uma região espaciotemporal variável indicando perigo. Desses termos referenciais originários mistos surgiriam, de um lado, termos singulares indexicais, de outro, termos gerais indexadores. Essa diferenciação, como as indicadas acima, não seria forçosa, mas, como notei, tendencial, uma vez que caminhos de derivação inversos e cruzados também são concebíveis.

---

<sup>4</sup> Para Tugendhat (1976, p. 380) o *quasi-predicado* é um predicado cuja regra de atribuição só lhe permite o emprego em situações de percepção apropriadas.

## Neodescritivismo aplicado a termos gerais

Não faria sentido defender uma teoria descritivista dos termos gerais para os termos gerais indexadores, posto que eles dizem respeito ao que podemos aqui qualificar como “simples”. Também não faria muito sentido tentar desenvolver semelhante teoria para o caso dos termos gerais descritivadores, posto que eles mesmos, quando funcionam propriamente, já podem ser traduzidos na forma de descrições indefinidas que exprimem suas próprias regras de atribuição. Contudo, podemos desenvolver explicações descritivistas para a aplicação de termos gerais nominadores.

Os termos de artefatos exemplificam de forma mais simples e transparente o último caso. Considere alguns exemplos de definições de semelhantes termos por intermédio de descrições indefinidas:

Cadeira (*Df.*) = um artefato não-veicular com encosto feito para uma só pessoa se sentar de cada vez.

Lápis (*Df.*) = um objeto manuseável, geralmente de madeira, com ponta de material suficientemente sólido e que serve para escrever.

Carro (*Df.*) = um objeto que se movimenta sobre rodas, sendo propriamente feito para transportar pessoas.

Catedral (*Df.*) = uma igreja que tem um trono de bispo e congrega as outras igrejas da diocese.

Podemos dizer que os termos acima abreviam as descrições *indefinidas* correspondentes, que, centralmente, exprimem seus significados cognitivos. Essas descrições constituem definições funcionais que expõem seus critérios de aplicação. Mas, diversamente do caso de espécies naturais, elas não fazem menção a essências subjacentes, posto que artefatos não as possuem. Podemos dizer que termos de artefatos têm o seu sentido determinado praticamente por completo por regras-descrições dos tipos que formam o conjunto <Dis + Disd> (descrições

indefinidas de superfície + descrições indefinidas disposicionais de superfície).

É verdade que definições descritivas dos termos nominadores de artefatos podem admitir e até requerer detalhamentos e que elas não possuem fronteiras precisas nem necessariamente imutáveis. Mas isso em nada as deprecia. A vaguidade é uma característica insuperável da grande maioria dos nossos conceitos, supostamente refletindo uma propriedade ontológica da realidade. O mais importante é que de posse (geralmente tácita) das definições, somos capazes de aplicar os conceitos a contento à maioria dos casos, o que lhes dá uma justificação pragmática. Pode-se tentar objetar contra as definições acima apresentadas apontando para casos limítrofes ou exceções aparentes. Um tronco de árvore com a forma de uma cadeira, que é trazido para casa e modelado de modo a ser usado como cadeira, é um caso limítrofe. Um lápis eletrônico, desses usados para se desenhar em telas de computador, pode ser parecido com um lápis, mas não é; trata-se já de um uso estendido da palavra, por analogia funcional.

Na identificação de um tipo psicólogos experimentais falam de *tipicalidade*, da aproximação de um dado particular com um *estereótipo* do particular. Assim, o estereótipo de lápis não é um lápis de cera, mas o velho lápis de madeira com ponta de grafite. O estereótipo de cadeira é a que encontramos na sala de jantar com quatro pernas e não a cadeira de balanço ou a de praia ou a cadeira elétrica. O estereótipo de um pássaro pode ser um gavião, mas não será um pinguim. Nós reconhecemos os estereótipos mais prontamente, já que eles apresentam propriedades mais típicas em maior número<sup>5</sup>. Isso torna fácil aos psicólogos serem induzidos à conclusão de que as maneiras tradicionais pelas quais os filósofos trataram conceitos, buscando *definições* com possíveis condições necessárias e suficientes de aplicação, é claramente incorreta.

Parece-me haver aqui uma séria confusão categorial. Uma coisa são as variáveis condições de *performance* que o psicólogo investiga; outra são os critérios derivados da regra definicional de atribuição da expressão

---

<sup>5</sup> Cf. Rosh (1973).

conceitual que interessam ao filósofo. As duas coisas são perfeitamente compatíveis, mas diversas. As estruturas lógico-conceptuais definitórias que constituem as regras de atribuição de uma palavra-conceito, as que constituem aquilo mesmo que podemos chamar de *conteúdo conceitual*, podem ser capazes de gerar múltiplas e variadas constelações criteriais que são responsáveis por subconceitos, variações semânticas mais e mais específicas. Assim, a regra de atribuição para o conceito de pássaro pode gerar constelações criteriais específicas, que são diferentes para passarinhos, águias, abutres, avestruzes, pinguins... entre passarinhos, constelações criteriais que nos permitam identificar espécies etc. Como é muito mais normal e frequente encontrarmos as constelações criteriais derivadas pelas quais reconhecemos pássaros que encontramos em árvores e que voam, somos mais rápidos em aplicá-la. Geneticamente, essas constelações devem ter vindo primeiro na formação do complexo criterial constitutivo da regra conceitual para a identificação de pássaros. O que os psicólogos experimentais descobriram foi apenas essa centralidade geradora de melhor performance na aplicação do predicado e não a razão originadora de sua aplicação.

Uma cadeira de balanço, por exemplo, não possui a mais frequente condição caracterizadora, que é a de ter quatro pés; mesmo assim, ela continua seguindo a definição conceitual de ser um artefato não-veicular com encosto, feito para uma só pessoa se sentar de cada vez<sup>6</sup>. Ela continua satisfazendo os critérios da regra de atribuição, aos quais não pertence a condição usual de ter quatro pés. E o mesmo vale para a cadeira de praia... Não é necessário, pois, que haja uma contradição entre uma definição filosófica tradicional (possivelmente apresentável em termos de condições necessárias e suficientes...) realizada em um nível mais abstrato, que explicita uma regra criterial que pode ser mais ou menos complexa, e a investigação de condições identificadoras frequentemente inessenciais,

---

<sup>6</sup> Se fosse veicular seria chamada de 'assento' (de carro, ônibus, avião...), se não tivesse encosto seria um banco, se tivesse sido feita para mais de uma pessoa se sentar poderia ser um sofá ou um banco de jardim... (Cf. Costa 2018, II, 7)

que, no dizer de Wittgenstein, não são critérios definitórios, mas apenas *sintomas* probabilizadores. Esses sintomas ou traduzem maior facilidade e rapidez no reconhecimento (como a cadeira de mesa ou um bem-te-vi) ou maior dificuldade e lentidão no reconhecimento (como a cadeira elétrica ou um pinguim).

Nos casos de espécies naturais como ‘água’ e ‘ouro’, parece que, como neodescritivistas, precisamos distinguir ao menos dois núcleos descritivos na constituição do sentido do termo: um núcleo popular, relacionado à nossa vida cotidiana, e um núcleo científico, relacionado à estrutura essencial subjacente, ambos os núcleos dependentes de uma regra disjuntiva inclusiva para a identificação do que possa ser chamado de ‘água’ ou ‘ouro’. Mas não parece possível generalizar essa duplicidade sequer para espécies naturais.

Consideremos o caso do conceito de água. Até o final do século XVIII o conceito de água era constituído por um núcleo popular. Tratava-se de descrições de superfície como as de um líquido transparente, inodoro, insípido, que enchia lagos, rios e mares, que possuía certas propriedades disposicionais, como a de saciar a sede e apagar o fogo... descrições que abrevio como Ds-H<sub>2</sub>O. Para o final daquele século, químicos começaram a descobrir a estrutura química da água, que após 1820 passou cada vez mais a firmar-se como H<sub>2</sub>O, junto ao que se poderiam adicionar propriedades como a de ser formada por uma reação produzida a partir de hidrogênio e água, simplificada:  $2H_2 + 2O = 2H_2O$ . No conjunto, tratava-se da descoberta do núcleo científico da água, que abreviamos como Dp-H<sub>2</sub>O. Com base nisso, podemos propor uma condição neodescritivista genérica para a identificação de água que demande tão somente uma disjunção dos dois núcleos de significado. Podemos agora analisar o conceito geral de água como sendo constituído por uma regra de caracterização que demanda a satisfação de uma disjunção inclusiva dos dois núcleos como se segue:

RA-‘água’:

O termo geral ‘água’ refere-se à propriedade particularizada (tropo) em qualquer instânciação de uma substância química em algum objeto  $x$  (identificável por um termo singular)

*see*

- (i)  $x$  satisfaz as regras de caracterização expressas pelo núcleo semântico popular Ds-água e/ou pelo núcleo semântico científico Dp-água associado à palavra e
- (ii) em medida no todo suficiente,
- (iii) mais do que a qualquer outra regra de atribuição.

Claro que há muito mais a se dizer sobre o conceito de água, como as *ênfases semânticas* que podemos fazer de um ou de outro núcleo na dependência do contexto, o que pode ser usado para se responder a conhecidos dilemas conceituas, mas não é aqui o lugar para entrar nesses detalhes.<sup>7</sup>

Considere um termo contável de espécie natural como ‘tigre’. Podemos considerar a descrição de superfície (o estereótipo) como sendo Dt = grande e feroz felino carnívoro e quadrúpede com pelo amarelado, listas escuras transversais e ventre branco. Junto a isso, o importante para a identificação de um tigre é que ele também pertença à espécie *Panthera tigris*, originária da Ásia, disso resultando Dte. Como resultado podemos chegar à seguinte regra metadescriptiva disposicional para a atribuição do conceito de tigre:

RA-‘tigre’:

Usamos o termo geral tigre para nos referirmos (propriamente) a uma propriedade de um animal  $x$

*see*

- (i) A propriedade de  $x$  satisfaz a regra de caracterização expressa por Dte,
- (ii) em medida suficiente,

---

<sup>7</sup> Ver Costa (2014): *Lines of Thought*, cap. III, On the Meaning of Water.

(iii) mais do que qualquer outra.

Ao menos para a zoologia atual, o termo de espécie natural ‘tigre’ tem funcionado de modo diferente do termo ‘água’. Enquanto a descrição da microestrutura é importante para a caracterização de massas de água, o *layout* genético não tem sido considerado essencial para a identificação de tigres, diversamente da constatação de sua pertinência a uma dada espécie. Pode ser que o conceito de tigre venha a ser alterado de modo que o *layout* genético passe a ter um papel mais relevante ou que a região originária da população deixe de ter relevância – regras conceituais não precisam ser fixas.

Ainda outro caso é o de termos gerais que ao menos parecem ser de espécie natural, mas que não possuem nenhuma estrutura subjacente. Considere conceitos geográficos, como os de rio e lagoa. Um rio é um fluxo de água suficientemente grande que corre sobre um leito naturalmente escavado na Terra em certa direção, partindo de uma nascente, diversamente de córregos, canais, rios de lava etc. Uma lagoa se distingue tão somente pela propriedade superficial de ser uma extensão fechada e suficientemente grande de água doce, diversamente de lagos, lagoas, poças de água etc. Uma alternativa diante desses casos seria a de redefinir o conceito de espécie natural demandando o compartilhamento pelos seus membros de uma essência subjacente, mesmo que descoberta *a posteriori*, o que com boas razões excluiria termos como os recém-considerados. No caso de espécies zoológicas como a dos tigres, é possível que a relação com o grupo historicamente e regionalmente localizado que deu origem ao nome possua alguma importância.

Há ainda casos em que além de serem múltiplos, os critérios caracterizadores do tipo de entidade se encontram muito fracamente conectados entre si. Nesses casos, podemos ser levados a recorrer a uma regra meta-caracterizadora para o termo geral capaz de estabelecer *o que e o quanto* da multiplicidade de condições que precisa ser aproximadamente satisfeito. Assim, podemos eventualmente precisar de

- (i) um conjunto de regras-descrições indefinidas de primeira ordem que constituem um *modelo* (o assim chamado feixe de descrições) e
- (ii) uma regra-descrição de ordem superior, RMD – a regra meta-descritiva do termo geral – estabelecendo o quanto do modelo precisa ser satisfeito para que o termo geral possa ser aplicado.

Exemplos explícitos de semelhantes regras criteriosais complexas demandadas para a aplicação de um termo conceitual são frequentes e totalmente explicitadas em ciência, como o demonstram o caso de uma doença como a febre reumática<sup>8</sup> ou de anormalidades como a síndrome de Asperger<sup>9</sup>. Tais regras demandam a satisfação de uma variável multiplicidade de condições criteriosais para poderem cumprir seu papel caracterizador.

Quero discutir aqui um único exemplo de conceito mais vago e aberto – aquilo que Wittgenstein classificaria algo equivocadamente como exprimível por palavras com ‘semelhanças de família’ (*Familienähnlichkeit*): o conceito de *religião*. Baseando-me em P. W. Alston, eis o modelo constituído das regras-descrições criteriosais mais relevantes para a aplicação referencial da palavra ‘religião’:

1. Crenças em seres sobrenaturais (deuses).
2. Sentimentos caracteristicamente religiosos (reverência, adoração, senso de mistério, sentimento de culpa etc. associados ao divino).
3. Um código moral que se acredita sancionado pelos deuses.
4. Rezas e outras formas de comunicação com os deuses.
5. Uma distinção entre entidades sagradas e profanas; atos rituais concentrados em torno de entidades ou objetos sagrados.
6. Uma ampla organização da vida individual e social baseada em características anteriormente descritas.
7. Uma cosmovisão, a saber, uma explicação do significado do mundo e do lugar do homem nele<sup>10</sup>.

---

<sup>8</sup> Cf. critérios de Jones atualizados.

<sup>9</sup> E.g. Gillberg (2002, cap. 2).

<sup>10</sup> Ver Alston (1964, cap. 6).

Como o próprio Alston notou, há religiões, como o catolicismo, o judaísmo e o islamismo, que satisfazem todas as regras-descrições constitutivas do feixe ou modelo. Elas constituem *casos prototípicos*, derivando-se a aplicação da palavra para casos que satisfazem apenas partes do modelo.

Cumpra notar que talvez nenhuma das condições incluídas no modelo seja necessária, disso resultando, ao que parece, a inexistência de uma essência geral única do que seja a religião em seu sentido lato. No protestantismo, nota Alston, os rituais relativos a objetos sagrados são bastante atenuados; e, no caso dos *Quakers*, eles são até mesmo repudiados, tornando-se fundamental apenas a experiência mística. Mesmo a crença em seres sobrenaturais pode estar ausente; há religiões como o Budismo Hinayana, nas quais os seres sobrenaturais são ignorados, incidindo a ênfase no cultivo de uma disciplina moral e meditativa que busque um estado espiritual em que todos os desejos parem de existir. E há religiões laicas em que a crença em um Deus pessoal sobrenatural é simplesmente excluída, como é o caso da Religião da Humanidade fundada por Auguste Comte, que diviniza a sociedade humana, além de possuir figuras devocionais (equivalentes aos santos) como Shakespeare e Clotilde de Vaux. Parece que, em uma extensão talvez indébita dos critérios sugeridos por Alston, a religião sequer precisa demandar uma prática social, como acontece com religiões pessoais, como a do filósofo Spinoza, que se baseava na calma e jubilosa aceitação de tudo o que acontecia como decorrente das leis impessoais do universo, ou a do físico Albert Einstein, que via nessas leis uma fonte de reverência e de encantamento.

Podemos com isso estabelecer uma regra de regras, uma metaregra aplicável às regras-descrições constitutivas do modelo formador do sentido do termo geral ‘religião’. Podemos chamá-la simplesmente de regra de atribuição constitutiva do conceito de religião ou:

RA-‘religião’:

Usamos a palavra 'religião' para referir-nos (propriamente) a uma prática sócio-cultural que é característica de um objeto *x*

*see*

- (i) *x* satisfaz suficientemente ao menos *uma* ou *duas* das regras-descrições constitutivas do modelo para o termo geral 'religião'.

Essa regra nos permite chamar de religião não só o catolicismo e o judaísmo, que exemplificam o próprio modelo, mas também o Budismo Hinayana e a Religião da Humanidade. Uma característica dessa regra de atribuição é que ela é propositadamente vaga. Ela é vaga de maneira a fazer sua aplicação corresponder à vaguidade de nosso próprio conceito de religião, que por sua vez deve corresponder à vaguidade intrínseca ao fenômeno considerado. Casos como o da religião da humanidade satisfazem a metaregra (condições 1, 5 e 6, pelo menos). Há casos de religiões pessoais (como as de Spinoza e Einstein) que são *fronteiriços*, no sentido de que quase não mais sabemos se o conceito realmente se aplica (satisfazem a condição 2). E há ainda casos que, eventualmente, aproximam-se de satisfazer alguma condição da regra de atribuição, embora não a satisfaçam suficientemente. Considere, por exemplo, organizações secretas como a dos Rosacruz, grupos políticos radicais como o dos comunistas ortodoxos, ou ainda, o misticismo matemático dos filósofos pitagóricos. Como as regras de classificação dos conceitos de organização secreta, grupos políticos radicais e filosofia se aplicam a essas coisas mais propriamente, o conceito de religião passa a aplicar-se a elas apenas em um sentido *estendido*.

### **Termos gerais e designação rígida**

Diante dessa variedade de regras de atribuição, podemos nos perguntar se dentro de concepções descritivistas mais sofisticadas, como a recém esboçada, os termos gerais poderiam ser considerados designadores rígidos. Tanto Kripke quanto Putnam, aderindo a uma concepção causal-externalista dos termos gerais, responderiam afirmativamente a essa pergunta, ao menos, certamente, no que concerne

aos termos de espécie natural, pois eles assumem que tais termos se referem a uma mesma essência microestrutural, ao menos em qualquer mundo no qual essa essência exista<sup>11</sup>.

Contudo, a rigidez dos termos gerais é diferente do caso menos problemático da rigidez dos nomes próprios. Enquanto o nome próprio designa apenas um único objeto em qualquer mundo possível no qual esse objeto exista, precisando-se apenas identificá-lo, o termo geral designa uma propriedade instanciável em um indefinido número de objetos que difere em cada mundo possível (um mundo possível pode não conter tigre algum, outro pode ser superpopulado por bilhões de tigres). Por isso mesmo, para quem defende a rigidez dos termos de espécie natural, a sua referência não deve poder ser a extensão. Também é problemático pensarmos que o termo geral é rígido por ter como referência uma espécie<sup>12</sup>. Pois se fizermos isso parece que, ou recaímos em problemas como o recém-exposto, ou então nós o circundamos concebendo a espécie como uma propriedade abstrata, mas ao preço de termos de aceitar um problemático realismo de propriedades.

Quero aqui coerentemente optar pela alternativa de que termos gerais designam primariamente propriedades espaço-temporalmente particularizadas, ou seja, *tropos*<sup>13</sup>. Sob esse ponto de vista, *um termo geral é rígido na medida em que ele designa uma mesma propriedade particularizada (tropo ou agrupamento de tropos) em qualquer mundo possível no qual essa propriedade particularizada exista*. Mas como cada mundo possível pode ter uma diversidade de propriedades particularizadas, precisamos de um critério de identidade para a propriedade a ser escolhida como sendo *a mesma* em cada mundo possível! Seria possível obter tal critério? Afinal, só os termos singulares parecem capazes disso, pois só eles são feitos para singularizar uma mesma coisa em qualquer mundo possível no qual ela exista.

---

<sup>11</sup> Kripke (1980, p. 134 ss.) e Putnam (1975, sec. 6).

<sup>12</sup> LaPorte (2000); ver também Schwartz (2002).

<sup>13</sup> Para acesso à ontologia dos tropos, ver Costa (2018, apêndice do capítulo III).

Para responder a essa questão, devo primeiro notar que sob a perspectiva da linguagem só podemos encontrar tropos como critérios para os termos gerais com auxílio da aplicação do critério de singularização oferecido por termos singulares, posto que eles são aquilo que nos diz onde, como, quando encontrá-los. Assim, na prática, será necessário vincularmos o termo geral a um termo singular e perguntarmo-nos se uma propriedade particularizada idêntica à do objeto referido pelo termo singular em nosso mundo existe ou, se não existe, se é, pelo menos, concebível. Nesse caso verificaremos que ela é rígida, pois em qualquer mundo possível no qual for identificado um objeto com uma propriedade particularizada qualitativamente idêntica a ela, essa propriedade será objeto de aplicação da regra de atribuição do termo conceitual correspondente.

Trata-se aqui, se quisermos, de uma *rigidez secundária*, dado que ela é dependente da rigidez dos termos singulares aos quais o termo geral vem associado em frases predicativas ou relacionais singulares. Considere, para exemplificar, o termo indexador ‘branco’. Quando usamos a frase predicativa singular “A Lua (da terra) é branca” estamos nos referindo à propriedade particularizada (tropo) da brancura da Lua, de modo que o termo geral indexador ‘branco’ pode ser considerado rígido por aplicar-se à mesma propriedade de brancura em qualquer mundo possível no qual, mesmo não existindo a Lua, existe algum objeto (entidade) qualquer com a propriedade de ser branco (analísável aqui em termos de reflexão de todos os comprimentos de onda do espectro sob certa perspectiva). A rigidez do predicado ‘...é branca’ é, em um sentido, secundária à rigidez de um nome próprio como ‘Lua’, o qual é rígido na independência de seu comparecimento em frases predicativas<sup>14</sup>.

Um outro caso é o de termos gerais nominadores como ‘ouro’. Suponhamos que a regra de atribuição para o que chamamos de ‘ouro’

---

<sup>14</sup> Com isso espero estar de acordo com a proposta de Devitt e Sterelny de que termos gerais são “aplicadores rígidos”: eles se aplicam a objetos no mundo atual, e se tais objetos existem em um mundo possível, então eles se aplicam a objetos nesse mundo possível. Devitt & Sterelny (1999, p. 85).

fosse expressa pela descrição “metal raro de cor propriamente amarelada, solúvel em água áurea, com peso atômico 76...”. Nesse caso, parece aceitável que através do proferimento indexical “Esse anel é de ouro”, feito por mim agora para me referir ao material de um anel, eu esteja usando o predicado como designador rígido, aplicando-o a um conjunto de propriedades qualitativamente idêntico em qualquer mundo possível no qual esse conjunto de propriedades seja individuado pela expressão ‘Esse anel’, que é um indexical complementado por um sortal, em circunstâncias similares.

Finalmente, para o caso dos termos gerais descritivadores isolados, considere a expressão predicativa “...tem senso de humor”. O predicado ‘ter senso de humor’ também parece ser um designador rígido secundário porque a propriedade em questão de ter senso de humor é singularizada em qualquer mundo possível no qual algum indivíduo compareça possuindo senso de humor.

Essa sugestão também explica porque termos gerais nominadores podem ser rígidos, enquanto as descrições constitutivas desses termos podem por contraste não ser rígidas. Assim, ‘água’ é um termo rígido por contraste com o termo descritivador ‘algo que serve para beber’, pois se aponto para esse copo de água e digo ‘Essa água serve para beber’ no mundo atual, isso é verdadeiro, mas pode ser que, em alguma circunstância contrafactual, esteja eu apontando para esse mesmo copo de água e esteja dizendo algo falso, pois, embora sendo a mesma água, ela contém algumas miligramas de arsênico e realmente não serve para beber. A explicação é aqui análoga a que poderia ser usada por um neodescritivista sofisticado para explicar a acidentalidade relativa das descrições definidas com relação aos nomes próprios<sup>15</sup>: o termo geral descritivador é, no caso, *semanticamente dependente* da aplicação do termo geral nominador, o que explica a sua flacidez relativa, do mesmo modo que a descrição definida é para o neodescritivista flácida por ser semanticamente associada de maneira frouxa às descrições constitutivas

---

<sup>15</sup> Cf. Costa (2014, cap. 2, p. 56-65).

do sentido do nome próprio que em nosso mundo teria um mesmo portador.

## Referências

- ALSTON, P. W. **Philosophy of Language**. Englewood Cliffs: Prentice Hall, 1964.
- BACKER, G. P. Criteria: A New Foundation for Semantics. **Ratio**, 16, 1974, p. 156-89.
- COSTA, Claudio. **Lines of Thought: Rethinking Philosophical Assumptions**. Newcastle upon Tyne: Cambridge Scholars Publishing, 2014.
- \_\_\_\_\_. **Philosophical Semantics: Reintegrating Theoretical Philosophy**. Newcastle upon Tyne: Cambridge Scholars Publishing, 2018.
- CRANE, Tim All the Difference in the World, **Philosophical Quarterly**, 41, 162, 1991, p. 1-26.
- DONNELLAN, Keith. Kripke and Putnam on Natural Kind Terms. In: C. GINET & S. SHOEMAKER (eds.). **Knowledge and Mind**. Oxford: Oxford University Press, 1983.
- DUPRÉ, John. Natural Kinds and Biological Taxa, **The Philosophical Review**, 90, 1981, p. 66-91.
- \_\_\_\_\_. **The Disorder of Things: Metaphysical Foundations of the Disunity of Science**. Cambridge MA: Harvard University Press, 1993.
- \_\_\_\_\_. Natural Kinds. In: W. H. Newton-Smith (ed.). **A Companion to the Philosophy of Science**. Oxford: Blackwell, 2000.
- ELBURNE, Paul. **Meaning: A slim guide to semantics**. Oxford: Oxford University Press, 2011.
- ERWIN, E., L. KLEIMAN, E. ZEMACH. The Historical theory of reference. **Australasian Journal of Philosophy**, 56, 1974, p. 50-6.
- HACKER, P. M. S. **Insight and Illusion**. Oxford: Oxford University Press, 1986.
- HEMPEL, C.G. **Philosophy of Natural Science**. Englewood Cliffs: Prentice-Hall, 1967.

KALLESTRUP, Jesper. **Semantic Externalism**. London: Routledge, 2012.

KRIPKE, Saul. **Naming and Necessity**. 2 ed. Cambridge: Harvard University Press, 1980.

KORNBLITH, Hilary. Referring to Artifacts, **Philosophical Review**, 89, 1980, p. 109-14.

LAPORTE, Joseph. Rigidity and Kind. **Philosophical Studies**, 97, 2000, p. 296-316.

\_\_\_\_\_. **Natural Kinds and Conceptual Change**. Cambridge: Cambridge University Press, 2004.

LEWIS, D. K. Naming the Colors. **Australasian Journal of Philosophy**, 75, 1997, p. 325-42.

LYCAN, William. The Meaning of Water: An Unsolved Problem. In: E. Sosa & E. VALLANUEVA (eds.). **Philosophy of Language, Philosophical Issues**, 16, 2006, p. 184-199.

MELLOR, D. H. Natural Kinds. **British Journal for the Philosophy of Science**, 28, 1977, p. 299-312.

PUTNAM, Hilary. **The Meaning of 'Meaning'**. Minneapolis: University of Minnesota Press, 1975, p. 131-93.

ROSCH, Eleanor. On the Internal Structure of Perceptual and Semantic Categories. In: MOORE, T. E. (ed.). **Cognitive Development and the Acquisition of Language**. New York: Academic Press, 1973.

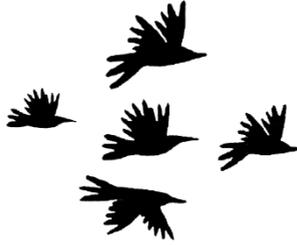
RUSSELL, Bertrand. The Philosophy of Logical Atomism. In: **Logic and Knowledge: Essays 1901-1950**. MARSCH, R. C. (ed.). London: Routledge, 1994 [1918].

SCHWARTZ, S. P. (ed.). **Naming, Necessity, and Natural Kinds**. Ithaca: Cornell University Press, 1977.

\_\_\_\_\_. Putnam on Artifacts. **Philosophical Review**, 87, 1978, p. 566-74.

\_\_\_\_\_. Kinds, General Terms, and Rigidity: A Reply to LaPorte. **Philosophical Studies**, 109, 2002, p. 265-77.

- \_\_\_\_\_. General Terms and Mass Terms. In: DEVITT, M. & HANLEY, R. (eds.). **The Blackwell Guide to Philosophy of Language**. Oxford: Blackwell, 2006.
- SMITH, A.D. Natural Kind Terms: A Neo-Lockean Theory. **European Journal of Philosophy**, 13, 2005, p. 70-88.
- STERELNY, Kim On Natural Kind Terms. **Pacific Philosophical Quarterly**, 44, 1983, p. 110-25.
- STROLL, Avrum. **Sketches and Landscapes: Philosophy by Example**. Cambridge MA: MIT-Press, 1996.
- TUGENDHAT, Ernst. **Vorlesungen zur Einführung in die sprachanalytische Philosophie**. Frankfurt: Suhrkamp, 1976.
- TUGENDHAT, Ernst & WOLF, Ursula. **Logische-Semantik Propädeutik**. Stuttgart: Reclam, 1983. (Trad. port.: TUGENDHAT, Ernst & WOLF, Ursula **Propedêutica Lógico-Semântica**. Petrópolis: Editora Vozes, 1997).
- WILLIAMS, D. C. The Elements of Being. **Review of Metaphysics**, vol. 7 (2), 3-18, 1953, p. 171-92.
- WITTGENSTEIN, Ludwig. **Philosophische Untersuchungen**. Frankfurt: Suhrkamp, 1984.
- ZEMACH, Eddy. Putnam's Theory on the Reference of the Substance Terms. **Journal of Philosophy**, 73, 1986, p. 116-27.



## Lewis contra Putnam, Fodor, Davidson e Kripke <sup>1</sup>

*Rodrigo A. dos S. Gouvea* <sup>2</sup>

### Introdução

Segundo o fisicismo, tudo o que existe é físico. Apesar de sua aparente simplicidade, há aspectos pouco claros nessa formulação. Para compreender e avaliar essa doutrina metafísica, precisamos saber o que é físico para o fisicismo<sup>3</sup>, assim como qual é o critério que emprega para reconhecer algo como existente. Ademais, caso reconheça a existência de entidades que não pareçam ser físicas, como crenças e sensações de dor, é preciso que se especifique o tipo de relação que vigora entre tais entidades e o que é físico. Neste trabalho, ocupo-me com a última demanda. Deve-se reconhecer que diferentes caracterizações do fisicismo se distinguem a partir de diferentes maneiras de conceber a relação entre o que é físico e o que não parece ser físico, mais especificamente, os estados mentais. Restrinjo-me aqui a investigar os prospectos de uma forma de conceber

---

<sup>1</sup> A pesquisa que originou este trabalho recebeu financiamento do Conselho Nacional de Desenvolvimento Científico (CNPq) e do Deutsche Akademischer Austauschdienst (DAAD).

<sup>2</sup> Professor adjunto no Departamento de Filosofia da Universidade Federal do Rio de Janeiro (UFRJ). Doutor em Filosofia pela Universidade de Leipzig (Alemanha). Atua como professor no PPGF da UFRJ e no PPGFIL da UFSJ. Coordenador do projeto *Mente e Representacionalismo*, financiado pelo CNPq através da Chamada MCTIC/CNPq N° 28/2018 - Universal.

<sup>3</sup> Para uma abordagem do problema de caracterização do que é físico para o fisicismo e uma tentativa de solução a partir da física ideal, ver Gouvea (2016).

essa relação, a saber, a tese que se convencionou chamar de “físicismo de tipos”.

A primeira caracterização do físicismo como uma doutrina metafísica surge da generalização da proposta de Ullin Place e J. J. C. Smart de que haveria uma relação de identidade entre sensações e processos cerebrais, uma relação de identidade entre tipos. Apesar de figurar como uma espécie de formulação original da referida doutrina metafísica, o físicismo de tipos foi duramente criticado. Caracterizações alternativas do físicismo foram oferecidas a partir das objeções que sugeriam problemas de formulação àquela. Entre as caracterizações alternativas, encontramos o físicismo de *tokens*, as teses físicistas baseadas em relações de superveniência, realização etc.<sup>4</sup> Contraindo-se a essa tendência, David Lewis defendeu o físicismo de tipos contra toda sorte de objeções. Para avaliar, ao menos parcialmente, os prospectos dessa caracterização da doutrina metafísica do físicismo, considerarei as objeções oferecidas por Hilary Putnam, Jerry Fodor, Donald Davidson e Saul Kripke ao físicismo de tipos, bem como as respostas que Lewis ofereceu ou poderia ter oferecido.

Após apresentar o físicismo de tipos nos moldes do que seria sua formulação original, descreverei as objeções a essa forma de físicismo que se baseiam nas seguintes considerações:

- i.) a tese da realizabilidade múltipla (Putnam e Fodor);
- ii.) a constatação do holismo do mental e do papel da racionalidade em sua estruturação (Davidson);
- iii.) a tese de que termos para sensações sejam designadores rígidos (Kripke).

Ao final, mostrarei como Lewis respondeu ou poderia responder a essas objeções. Em razão de exercer um papel fundamental na posição de Lewis, o funcionalismo analítico a respeito da mente será elucidado. Em seguida, argumentarei que:

---

<sup>4</sup> Um panorama crítico das teses físicistas baseadas em relações de superveniência e de realização é oferecido por Gouvea (2020).

- i.) Lewis soluciona o problema da realizabilidade múltipla por meio de identidades específicas a espécies;
- ii.) a natureza holística do mental e sua ordenação racional são reconhecidas e tornadas não-problemáticas em decorrência das definições funcionais para termos mentais serem obtidas, como um pacote completo, a partir do postulado da psicologia popular;
- iii.) a identidade entre tipos mentais e físicos é contingente, em razão de os termos para estados mentais não serem designadores rígidos.

## 1. A teoria da identidade e o fisicismo de tipos

A afirmação de que sensações são processos cerebrais consiste na tese central da teoria da identidade de Place (1956) e Smart (1959). Em sua formulação original, a teoria da identidade se apresentava como uma alternativa tanto ao dualismo metafísico quanto ao fisicismo semântico<sup>5</sup>.

Uma teoria dualista potencialmente interessante, que é confrontada por Place e Smart, afirma que estados mentais são correlacionados, mas não são idênticos a estados físicos. A crítica que os referidos autores fazem a essa forma de dualismo não é direta. Consiste fundamentalmente em uma defesa da teoria da identidade como uma alternativa aceitável e mais apropriada, em vista de princípios de adequação empírica e simplicidade. Place (1956) descreve sua posição como uma “hipótese científica razoável, do modo como ‘luz é o movimento de cargas elétricas’ é uma hipótese científica razoável” (Place, 1956, p. 45). Smart (1959) evita essa caracterização e enfatiza que o debate deva ser travado em solo filosófico. No referido artigo, no entanto, Smart não oferece um argumento

---

<sup>5</sup> Concebido como uma tese metafísica, o fisicismo gozou de grande influência em debates filosóficos na segunda metade do século XX. Em seu uso original, no entanto, o termo “fiscicismo” dava nome a uma tese semântica. O termo foi cunhado por Otto Neurath para se referir à posição de que a física promoveria a unidade das ciências (1931a, p. 2 & 1931b, p. 620), no sentido de que “toda afirmação [com sentido] inclui determinações a respeito da ordenação espaço-temporal, ordenação que conhecemos através da física” (1981 [1931], p. 419). Neurath enfatiza a distinção entre sua concepção de fisicismo e a velha doutrina do materialismo. Diz que “[o] fisicismo não defende a tese de que o ‘mental’ (*das Geistige*) seja um produto da ‘matéria’, mas sim que tudo acerca do qual se possa falar com sentido são ordenações espaço-temporais” (1931a, p. 11). Rudolf Carnap desenvolveu essa concepção semântica. Segundo Carnap (1932), o fisicismo se caracteriza pelas seguintes alegações: (i.) “a linguagem física é uma linguagem universal, i.e., uma linguagem a que todas as sentenças podem ser traduzidas”; e (ii.) “para cada conceito psicológico (i.e., expressão) uma definição pode ser disposta, através da qual ele seja redutível de modo imediato ou mediado a conceitos físicos” (p. 107 & p. 109).

significativo em favor da teoria da identidade e admite que sua defesa é “uma confissão de fé” sustentada somente pelos “princípios de parcimônia e simplicidade” (Smart, 1956, p. 143 & p. 156).

Em contraste com a identidade expressa por “Phosphorus é Phosphorus”, a identidade entre sensações e processos cerebrais foi caracterizada pelos proponentes da teoria da identidade como contingente. Segundo Place, essa relação de identidade seria similar às identidades científicas, as quais afirma que não sejam necessariamente verdadeiras nem necessariamente falsas (Place, 1956, p. 45). Smart afirma de forma mais direta que a identidade entre “processos cerebrais e experiência [consciente] seja contingente” (Smart, 1959, p. 152). De acordo com Smart, identidades psicofísicas se assemelhariam à identidade contingente expressa por “a Estrela da Tarde e a Estrela da Manhã são uma e a mesma coisa” (Smart, 1959, p. 147s.).

Há diversas razões para assumir que relações de identidade entre sensações e processos cerebrais sejam contingentes. Como no caso de “Phosphorus é Hesperus”, a verdade do proferimento de “dor é estimulação da fibra-C” só pode ser averiguada *a posteriori* (Smart, 1959, p. 146s.). Além disso, os termos que empregamos em sentenças que exprimam relações de identidade psicofísicas têm significados distintos, apesar de poderem se referir à mesma coisa (Smart, 1959, p. 147s; Place, 1956, p. 146s.). Por fim, como indica Saul Kripke, temos a impressão de que nesses casos as coisas poderiam ter sido diferentes. Podemos imaginar que dor, p. ex., possa ser instanciada em “criaturas totalmente desincorporadas (*disembodied*)”, ou que os processos cerebrais que sejam correlatos a ocorrências de dor em seres humanos, digamos, estimulações da fibra-C, possam ocorrer sem que haja instanciações de dor (Kripke, 2012 [1971], p. 136).

O fisicismo de tipos foi tradicionalmente compreendido como uma generalização da teoria da identidade de Place e Smart. Restringindo-se à filosofia da mente, o fisicismo de tipos é a tese de que há identidade entre

os tipos de entidades mentais e tipos de entidades físicas<sup>6</sup>. Há um risco de equivocidade acerca de quais sejam os *relata* das identidades propostas pelo fisicismo de tipos, dado que a noção de tipos foi caracterizada de diferentes maneiras. O significado de “tipo” já foi elucidado por apelo a espécies, universais, conjuntos e outras coisas (Wetzel 2006). No entanto, interpretações equívocas a respeito de quais sejam os *relata* das identidades propostas pelo fisicismo de tipos devem ser evitadas com base nas observações de Smart e Place de que tais identidades concernem os referentes dos termos. Smart afirma que, em proferimentos de identidades psicofísicas, como em “experiência é processo cerebral”, ele não está afirmando que as expressões empregadas “... signifiquem o mesmo, ou até que tenham a mesma lógica” (Smart, 1959, p. 151). Place também reconhece “a independência lógica de duas expressões” cujos referentes não são, de acordo com sua perspectiva, ontologicamente independentes (Place, 1956, p. 46). A identidade de tipos é concebida, assim, como uma relação de identidade entre os referentes dos termos de um proferimento de identidade. Tais referentes são *tokens*, ou ocorrências concretas dos tipos em questão. Desse modo, ao menos de acordo com o fisicismo de tipos, uma identidade de tipos deve ser compreendida como uma relação de identidade entre todos os *tokens* de um certo tipo com todos os *tokens* de outro tipo. Um exemplo ilustrativo pode ser expresso por “dor é estimulação da fibra-C”. Ao proferir essa sentença, afirmamos que todos os *tokens* de dor, i.e., todas as ocorrências ou instâncias de dor, sejam *tokens* de estimulação da fibra-C. A afirmação de “água é H<sub>2</sub>O” seria outro exemplo de uma identidade de tipos, a qual diz que todas as instâncias de água sejam instâncias do tipo H<sub>2</sub>O.

---

<sup>6</sup> O termo “entidade” está sendo empregado aqui de forma a poder abranger elementos de diferentes categorias ontológicas, como a dos objetos, das propriedades, dos eventos etc.

## 2. Objeções de formulação ao fisicismo de tipos

Esta seção apresentará objeções ao fisicismo de tipos que visam demonstrar sua implausibilidade com base no modo como foi formulado. Como indicado pelo título deste trabalho, as objeções abordadas aqui foram propostas por Putnam, Fodor, Davidson e Kripke.

### 2.1. Putnam e Fodor e o argumento da realizabilidade múltipla

Por propor que todos os *tokens* de um determinado tipo mental sejam *tokens* de um certo tipo físico, o fisicismo de tipos é uma tese controversa. É incompatível, por exemplo, com a constatação de que dois *tokens* de um dado estado mental possam estar correlacionados a *tokens* de diferentes tipos físicos. Um famoso argumento contra o fisicismo de tipos, a saber, o argumento da realizabilidade múltipla, enfatiza a incompatibilidade da tese em questão com considerações bastante plausíveis acerca do modo como o mundo seja ou possa ser. Em poucas palavras, o argumento da realizabilidade múltipla afirma ser muito plausível que *tokens* de um tipo mental sejam instanciados quando *tokens* de diferentes tipos físicos forem instanciados. Hilary Putnam defendeu originalmente essa posição. Ao usar dor como exemplo, caracterizou como “uma hipótese ambiciosa” que *tokens* de dor em mamíferos, répteis e polvos ocorram somente quando *tokens* de um mesmo tipo físico forem instanciados (1975 [1967], p. 436s.).

Jerry Fodor também argumentou de forma influente em favor da possibilidade de múltipla realização de tipos mentais, mas, como Putnam, não ofereceu um argumento muito enfático. A abordagem de Fodor se distingue por estender a tese da realizabilidade múltipla para além da psicologia. Diz respeito, em sua generalidade, aos tipos de níveis mais elevados (*higher-level types*) referidos pelas diversas ciências especiais, i.e., aquelas que não são a física. Sua defesa da implausibilidade das identidades de tipos propostas pelo fisicismo de tipos se revela mais

substancial que aquela proposta por Putnam, compreendendo as seguintes considerações:

(a) generalizações interessantes (...) podem ser frequentemente feitas acerca de eventos cujas descrições físicas não tenham nada em comum; (b) é frequentemente o caso que se as descrições físicas desses eventos subsumidos por tais generalizações tenham qualquer coisa em comum, isso seja (...) inteiramente irrelevante para a verdade das generalizações (...); (c) o trabalho das ciências especiais consiste em grande medida em formular generalizações dessa espécie (Fodor, 1980 [1974], p. 124).

## **2.2. Davidson e a objeção baseada no holismo do mental e sua ordenação racional**

Uma crítica ao fisicismo de tipos que possui similaridades com o argumento da realizabilidade múltipla se segue de uma interpretação da referida tese. Ao examinar a teoria da identidade, Donald Davidson afirma que “se as únicas identidades são de tipos de eventos, a teoria da identidade pressupõe leis de correlação” (1980 [1970], p. 110). Ele desenvolve, a partir dessa constatação, uma objeção ao fisicismo de tipos ao argumentar que o mental seja “nomologicamente irreduzível” ao físico (Davidson, 1980 [1970], p. 112).

A relação anômala entre o mental e o físico é fundada no suposto holismo do mental (Davidson, 1980 [1970], p. 113 & 1994, p. 231), o qual consiste na relação de interdependência que estados mentais mantêm uns com os outros. Davidson a caracteriza em termos mais específicos da seguinte forma:

Atitudes proposicionais, em termos das quais eventos mentais são caracterizados, não podem existir isoladamente. Crenças, intenções, dúvidas e desejos individuais devem suas identidades em parte à sua posição em uma grande rede de outras atitudes; o caráter de uma dada crença depende de infinitas outras crenças; crenças têm o papel que têm por causa de suas relações com desejos e intenções e percepções. Essas relações entre as atitudes são essencialmente lógicas: o conteúdo de uma atitude não pode ser divorciado

do que implica (*what it entails*) e do que é implicado por ele (*what is entailed by it*)<sup>7</sup> (Davidson, 1994, p. 231s.).

Dois aspectos da interdependência de estados mentais são descritos na passagem citada: a posse de um estado mental é dita depender da posse de outros estados mentais; e a própria identidade de um estado mental dependeria *em parte* de outros estados mentais a que o primeiro está logicamente relacionado. Para ilustrar ambos os pontos, considere a crença de que haja cerveja na geladeira. Uma pessoa não pode entreter esse estado mental isoladamente. Ela precisa acreditar, por exemplo, que ao menos uma cerveja e uma geladeira existam; que recipientes que guardem cervejas caibam na geladeira em questão, que geladeiras tem a função de resfriar etc. Essas crenças são implicadas (*entailed*) pelo conteúdo da crença de que haja cerveja na geladeira. A identidade ou o caráter de um estado mental depende em parte da multiplicidade de estados mentais que ele implique, visto que sem eles, no caso em que não sejam instanciados, o primeiro não poderia ser efetivamente entretido.

De acordo com Davidson, reconhecer a relação de interdependência entre estados mentais não é o bastante para justificar sua irredutibilidade ao físico, dado que uma relação de interdependência também deve vigorar entre as entidades físicas<sup>8</sup>. Ao invés, essa espécie de anomalia psicofísica se segue dos “compromissos díspares” que regulam os dois domínios (Davidson, 1980 [1970], p. 116). Entre estados mentais, espera-se que a racionalidade regule suas ocorrências de acordo com relações essencialmente lógicas que as conecte. Se um estado mental é instanciado, como a crença de que há cerveja na geladeira, então ele deve ser acompanhado por estados mentais que implica. O fato de que a racionalidade não executa esse papel no domínio físico revela uma

---

<sup>7</sup> É estranho que Davidson caracterize os estados mentais a que um dado estado mental seja interdependente por meio das expressões “o que implica (*what it entails*) e o que é implicado por ele (*what is entailed by it*)”, uma vez que se referem às mesmas coisas. O que algo implica simplesmente é aquilo que é implicado por ele.

<sup>8</sup> A interdependência de entidades físicas é defendida com base no “aspecto da realidade física de que uma mudança física possa ser explicada por leis que a conectem com outras mudanças e condições fisicamente descritas” (Davidson, 1980 [1970], p. 116).

diferença que justifica, segundo Davidson, a irredutibilidade nomológica do mental ao físico e a falsidade do fisicismo de tipos.

### **2.3. Kripke e a objeção a partir designação rígida de termos empregados em identidades psicofísicas**

A objeção de Kripke ao fisicismo de tipos enfatiza a dificuldade em determinar a natureza modal das alegadas relações de identidade entre estados de consciência e estados físicos<sup>9</sup>. Em primeiro lugar, Kripke revela que se um proferimento de identidade dessa espécie é verdadeiro, ele deve ser necessariamente verdadeiro. Assim, não pode expressar uma identidade contingente como Smart e Place originalmente sugeriram. Em segundo lugar, Kripke ressalta as dificuldades de executar a tarefa que se impõe à proponente da teoria da identidade, ou do fisicismo de tipos, se reconhecer tais identidades como necessárias. Há razões para considerar relações psicofísicas de identidade como sendo contingentes. Uma dessas razões é a aparência de contingência, i.e., a aparência de que as coisas poderiam ter sido diferentes. De acordo com Kripke, se a proponente do fisicismo de tipos reconhecer que identidades psicofísicas sejam necessárias, ela terá a tarefa de explicar de modo a afastar (*explain away*) a aparente possibilidade de que as coisas poderiam ter sido diferentes. Em outras palavras, ela terá que mostrar que a aparente contingência não é possível e nem sequer imaginável (Kripke, 2012 [1971], p. 137). Kripke registra seu pessimismo sob a forma de um eufemismo, ao ressaltar que a tarefa de explicar de modo a afastar (*explain away*) a aparência de contingência demandaria argumentos muito sutis, que talvez não possam ser obtidos.

Antes de argumentar contra a tese de que proferimentos de identidades psicofísicas sejam contingentes, Kripke defende que os casos de proferimentos de identidades que empreguem apenas nomes próprios

---

<sup>9</sup> Veja a nota 18 de *Identity and necessity* (Kripke, 2012 [1971], p. 139s) para uma breve caracterização do argumento de Kripke contra o fisicismo de tipos.

(referidos, a seguir, por “nomes”), assim como os casos de identificações científicas/teóricas, são necessariamente verdadeiros ou necessariamente falsos. Esses são casos que Smart e Place descrevem como sendo análogos a identificações psicofísicas, no sentido de que todos exprimiriam identidades contingentes. A posição de Kripke se baseia na concepção de que nomes, como “Muhammad Ali” e “Cassius Marcellus Clay Jr.”, e os termos para tipos naturais (*natural kind terms*) empregados em identificações científicas/teóricas, sejam designadores rígidos. Nas palavras de Kripke, um designador rígido é “um termo que designa um mesmo objeto em todos os mundos possíveis” (Kripke, 2012 [1971], p. 127)<sup>10</sup>. A designação ou referência é rigidificada, por assim dizer, de acordo com o modo como nós usamos esses termos no mundo atual. Portanto, segundo Kripke, mesmo em um mundo possível em que Cassius Clay não tenha mudado seu nome, o nome “Muhammed Ali” ainda se refere a ele<sup>11</sup>; e, em um mundo possível em que seres humanos sejam cegos, o termo “luz” ainda irá se referir ao movimento de cargas elétricas.

A partir da caracterização de nomes como designadores rígidos, se um proferimento de identidade que empregue apenas nomes for verdadeiro, ele será necessariamente verdadeiro, verdadeiro em todos os mundos possíveis (Kripke, 2012 [1971], p. 132 & 1981 [1972], p. 102). Considere, e.g., “Hesperus é Phosphorus”. Trata-se de uma sentença que emprega apenas nomes, os quais atuariam como designadores rígidos, referindo-se às mesmas coisas em todos os mundos possíveis. Dado que seu proferimento é verdadeiro, não é apenas no mundo atual que “Hesperus” e “Phosphorus” se referem ao mesmo objeto, a saber, o planeta Vênus, mas sim em todos os mundos possíveis que o contenham, o que torna o proferimento uma verdade necessária. O mesmo se aplica aos casos de identificações científicas/teóricas (Kripke, 2012 [1971], p. 135 &

---

<sup>10</sup> Em contraste, um designador não rígido é um termo que designa, sem qualquer mudança no modo com que o aplicamos, diferentes objetos em cenários contrafactuais (Kripke, 2012 [1971], p. 127). Segundo Kripke, descrições são tipicamente designadores não rígidos.

<sup>11</sup> É importante notar que, ao modificar seu nome, “Muhammad Ali” passou a designar o grande boxeador também em estágios anteriores de sua vida. Muhammad Ali não nasceu em um cartório em 1964, mas no Hospital Geral de Louisville em 1942.

1981 [1972], p. 140). A partir da posição de que os termos “luz” e “movimento de cargas elétricas” se referem às mesmas coisas em todos os mundos possíveis, se o proferimento da sentença “luz é movimento de cargas elétricas” for verdadeiro em nosso mundo, então será verdadeiro em todos os mundos possíveis.

Como dito acima, podemos reconhecer dois passos na crítica de Kripke à teoria da identidade. Primeiramente, ressalta que a alegada identidade entre estados de consciência e entidades físicas não pode ser contingente. Essa conclusão se segue da tese de que os termos com os quais nós nos referimos a tais entidades sejam também designadores rígidos (Kripke, 2012 [1971], p. 136). Kripke constrói seu argumento com base no exemplo “dor é estimulação da fibra-C”. O termo “dor” é caracterizado como um designador rígido, porque se algo é dor, i.e., se algo é sentido como dor, então é necessariamente dor (Kripke, 1981 [1972], p. 149). Assim, “dor” se refere a instâncias do tipo dor em todos os mundos possíveis em que dor for instanciada. O termo “estimulação da fibra-C” é considerado como um designador rígido, dado que, supostamente, é um termo para tipo natural (*a natural kind term*). Caso não seja, Kripke nos diz, podemos simplesmente substituí-lo por um termo que designe rigidamente o processo cerebral usualmente correlacionado à dor (Kripke, 1981 [1972], p. 149). A partir da posição de que “dor” e “estimulação da fibra-C” são designadores rígidos, o proferimento “dor é estimulação da fibra-C” deve ser necessariamente verdadeiro, caso seja verdadeiro no mundo atual.

O segundo passo da objeção de Kripke à teoria da identidade – e ao fisicismo de tipos que, como vimos, é uma generalização daquela – consiste em enfatizar a dificuldade de assumir a alegada identidade entre estados de consciência e entidades físicas como sendo necessária. A tese de que identidades psicofísicas sejam necessárias contrasta com a aparência de sua contingência, i.e., o fato de que podemos conceber ou imaginar as coisas como sendo de outra forma. Caso pretendamos reconhecer que a identidade entre estados de consciência e processos cerebrais seja

necessária, restaria a tarefa de explicar de modo a afastar (*explain away*) a aparência de contingência. De acordo com Kripke, essa tarefa não poderá ser satisfatoriamente executada.

Kripke revela uma forma de explicar e rejeitar a aparência de contingência de proferimentos de identidades científicas/teóricas e de identidades que contenham apenas nomes. Afirma, no entanto, que a contingência aparente de identificações entre estados de consciência não pode ser explicada e rejeitada através da aplicação do mesmo procedimento. Apesar das semelhanças, os primeiros casos diferem dos últimos em um aspecto importante. Diferem no modo como podem ser reconhecidos os referentes dos termos que figuram na relação da identidade. No caso dos nomes e dos termos para tipos naturais, há propriedades contingentes que podem atuar como componentes intermediários em nosso reconhecimento de seus referentes. Nós podemos reconhecer o referente de um nome por meio de uma propriedade contingente que o referente manifeste com exclusividade. No caso de “Muhammad Ali”, por exemplo, podemos nos servir da propriedade apresentada por “x é o boxeador profissional estadunidense que mudou seu nome para ‘Muhammad Ali’ após se converter ao islamismo”. De modo similar, podemos reconhecer os referentes de um termo para tipo natural, como “luz”, por meio de uma propriedade contingente que exclusivamente manifeste, como a propriedade de produzir sensações de cores em seres humanos. Apesar de nos permitirem identificar os referentes de “Muhammad Ali” e “luz” no mundo atual, ambas as propriedades são contingentes. Elas não caracterizam os referentes desses designadores rígidos em todos os mundos possíveis que contenham tais referentes. Por exemplo, em um mundo possível em que Cassius Marcellus Clay Jr. não mudou seu nome, a propriedade mencionada acima não o caracteriza, apesar de o termo “Muhammad Ali” lhe designar.

De acordo com Kripke, não há uma propriedade contingente que execute o papel de intermediação no reconhecimento de instâncias de

estados de consciência. Baseando sua posição nos casos paradigmáticos de dor, Kripke argumenta que não há uma descrição e nenhuma propriedade contingente por meio da qual nós identificamos uma instancia de dor. Defende que seja “selecionada (*picked out*) pela propriedade de ser ela própria dor, por sua qualidade fenomenológica imediata” (Kripke, 1981 [1972], p. 152). A propriedade de ser sentida como dor é uma propriedade essencial da dor. Diferentemente da propriedade de produzir sensações de cores, no caso da luz, e de sensações de calor, no caso do calor, à dor não pode faltar sua qualidade fenomenológica. Se essa propriedade não for manifestada, não há instanciação de dor. Kripke propõe que todos os termos que se refiram a estados de consciência tenham seus referentes identificados por propriedades essenciais similares. Assim, quem defenda a teoria da identidade ou o fisicismo de tipos não pode apelar a propriedades intermediárias para explicar de modo a afastar a aparente contingência de identidade psicofísicas, como a que expressamos em “dor é estimulação da fibra-C”. Dado que não parece haver nenhum procedimento alternativo para explicar a aparente contingência como ilusória, temos razão para assumir que o proferimento da sentença “dor é estimulação da Fibra-C” não seja necessariamente verdadeiro. No entanto, isso falsifica o consequente do condicional apresentado acima, a saber, o de que tais identidades sejam necessariamente verdadeiras se forem verdadeiras no mundo atual, uma vez que se trata de uma identidade entre referentes de designadores rígidos. Kripke conclui que se “dor é estimulação da fibra-C” e outros proferimentos de identidades entre estados de consciência e entidades físicas não são necessariamente verdadeiros, então não são verdadeiros no mundo atual e o fisicismo de tipos é falso.

O argumento em questão pode ser apresentado esquematicamente do seguinte modo:

1. Dado que “dor” e “estimulação da fibra-C” são designadores rígidos, então, se o proferimento de “dor é estimulação da fibra-C” for verdadeiro, será necessariamente verdadeiro.

2. Temos razões para assumir que o proferimento de “dor é estimulação da fibra-C” não seja necessariamente verdadeiro.

C. Temos razões para assumir que o proferimento de “dor é estimulação da fibra-C” e o fisicismo de tipos sejam falsos.

### 3. Uma defesa lewisiana do fisicismo de tipos

A tese denominada de funcionalismo analítico tem lugar central na caracterização e defesa que David Lewis oferece ao fisicismo de tipos. Dessa concepção de estados mentais, pode-se extrair respostas às objeções de Putnam, Fodor, Davidson e Kripke apresentadas acima. Em suas diferentes versões, o funcionalismo acerca da mente professa que a natureza de um estado mental é caracterizada por seu papel funcional. Um papel funcional dessa espécie é o conjunto de causas e efeitos típicos do estado mental juntamente com as relações que mantém com outros estados mentais. Deve-se atentar que, segundo *o funcionalismo analítico* de Lewis, um estado mental é idêntico àquilo que executa seu papel funcional, e nós podemos especificar o papel funcional de todos os estados mentais ao analisar a psicologia popular.

Lewis (1970 & 1999 [1972]) argumenta que o significado de um termo que seja introduzido por uma teoria é determinado pelo uso que fazemos do termo no postulado da teoria. De acordo com Lewis, uma teoria que introduz termos (doravante, T) é formulada por um postulado, o qual inclui termos que são introduzidos pela teoria, os termos-T, e alguns termos antigos, cujo significado nos é familiar, os termos-O (Lewis, 1970, p 428s.). O postulado de T é uma sentença – provavelmente uma conjunção de sentenças – que apresenta relações causais e de outras espécies entre os referentes daqueles termos. Em uma caracterização que enfatiza a presença dos novos termos-T, o postulado é formalizado como:

$$T[t_1 \dots t_n]$$

Apesar de reconhecer que seja um mito, Lewis nos pede para tomarmos a psicologia popular (doravante, TF) como uma teoria que introduz termos novos. Seu postulado seria composto por uma conjunção de princípios da psicologia popular, os quais podemos tornar explícitos através da análise conceitual. A psicologia popular é caracterizada como apresentando relações entre tipos de estados mentais, condições ambientais que frequentemente atuam como estímulos e formas de comportamento que se espera que sejam causadas por ocorrências dos estados mentais (Lewis, 1999 [1994], p. 299). Em outras palavras, apresenta relações entre tipos de estados mentais, suas causas e seus efeitos típicos. Lewis argumenta que essas relações, assim como as regras gramaticais, são usualmente conhecidas de forma implícita. Por meio da análise conceitual, poderíamos torná-las explícitas.

O postulado de TF, que tornaríamos explícito através da análise conceitual, contém termos para estados mentais (*mental state terms*), os termos-T da psicologia popular e termos que não denotam estados mentais, seus termos-O, como os termos físicos e quase-lógicos. O postulado da suposta teoria da psicologia popular pode ser escrito formalmente de um modo que enfatize a presença dos termos para estados mentais:

$$TF[m_1...m_n]$$

Se substituirmos os termos para estados mentais no postulado de TF por variáveis, e prefixarmos a fórmula com quantificadores existenciais, obteremos uma sentença de Ramsey de TF:

$$\exists x_1...x_n TF[x_1...x_n]$$

A sentença de Carnap de TF é escrita formalmente do seguinte modo:

$$\exists x_1...x_n TF[x_1...x_n] \rightarrow TF[m_1...m_n]$$

Primeiramente, a sentença de Carnap afirma que se a sentença de Ramsey for verdadeira, então TF será verdadeira. Entretanto, ela também sugere algo relevante acerca do significado dos termos para estados mentais empregados em TF. Sugere que se a sentença de Ramsey de TF for verdadeira, então devemos compreender o significado dos termos para estados mentais de um modo que torne TF verdadeira.

A verdade de uma sentença de Ramsey de uma dada teoria depende de se há um ênuplo de entidades que execute os papéis indicados pelas variáveis. A teoria como um todo (apresentada como consequente do condicional que configura a sentença de Carnap) é verdadeira se os termos que introduz, seus termos-T, referem-se a um ênuplo de entidades que a realize. Lewis defende que os termos-T de uma teoria devem se referir apenas às entidades que realizam a teoria, caso elas existam. Desse modo, a sentença de Carnap oferece um ponto de partida para interpretar os termos-T de uma teoria. A respeito de TF, podemos caracterizar o significado dos termos para estados mentais do seguinte modo:

$$m_1...m_n = X_1...X_n TF[X_1...X_n]$$

Essa é a definição funcional oferecida por Lewis para os termos para estados mentais.

De acordo com Lewis, um estado mental é o estado físico que executa seu papel funcional. Ele oferece o seguinte esquema em favor dessas identificações:

estado mental M = o ocupante do papel-m (por análise)

estado físico P = o ocupante do papel-m (pela ciência)

portanto, M = P (Lewis, 1999 [1994], p. 303 & 1999 [1972], p. 248s.).

Ao analisar a psicologia popular por meio do procedimento apresentado acima, podemos especificar a natureza de um estado mental como aquilo que ocupa um certo papel funcional. Investigações empíricas de caráter científico podem nos levar a descobrir qual estado físico executa ou ocupa o papel funcional em questão. Desse modo, a caracterização

funcional do estado mental e a identificação de um estado físico como o ocupante do papel funcional característico do estado mental justificam um proferimento de identidade entre o estado mental e o estado físico.

A partir do funcionalismo do mental, Lewis e outros ofereceram caracterizações mais convincentes do fisicismo de tipos do que aquelas que obtemos através da simples generalização da teoria da identidade de Place e Smart<sup>12</sup>. A primeira defesa que Lewis faz do fisicismo de tipos é apresentada em “An Argument for the Identity Theory” (1966). O funcionalismo analítico do mental figura como a primeira premissa de seu argumento (1966, p. 19s.).

A segunda premissa do argumento em questão é a hipótese de que “um corpo unificado de teorias científicas... [possa] oferecer um relato (*account*) verdadeiro e exaustivo dos fenômenos físicos” (Lewis, 1966, p. 23). Lewis argumenta que essa seja uma hipótese de trabalho tradicional e definitiva da ciência natural” e enfatiza que não está em relação de implicação com o fisicismo (1966, p. 23). Com efeito, a hipótese não declara a inexistência de entidades não físicas, mas apenas que “fenômenos físicos sejam fisicamente explicáveis” (Lewis, 1966, p. 23s.). A partir dessa hipótese, também referida por outros autores como “o princípio do fechamento causal do mundo físico”, Lewis extrai uma tese acerca da natureza das entidades não físicas. Ele argumenta que se tais coisas existem, elas não podem ser causa de fenômenos físicos (Lewis, 1966, p. 24).

O argumento, que denominarei de “argumento causal de Lewis”, funciona do seguinte modo. A primeira premissa, uma afirmação do funcionalismo analítico acerca da mente, identifica estados mentais com as entidades que executam seus papéis funcionais. Papéis funcionais são conjuntos de causas e efeitos típicos juntamente com relações entre

---

<sup>12</sup> Lewis reconhece que David Armstrong propôs, de forma independente, uma formulação do fisicismo de tipos que se assemelha à sua, ao defender identidades psicofísicas de tipos com base em uma concepção funcionalista de estados mentais (Lewis, 1983 [1980], p. 124 & 1999 [1994], p. 303). Jackson, Pargetter e Prior (1982) também argumentaram de modo influente em favor do fisicismo de tipos com base no funcionalismo. Entretanto, a tese que defendem se difere daquela apresentada por Lewis e Armstrong em alguns aspectos. Uma defesa recente do fisicismo de tipos na linha de Lewis e Armstrong é oferecida por Jackson (2012).

estados mentais, os quais são revelados por análise da psicologia popular. Entre os efeitos típicos de estados mentais, reconhecemos diferentes manifestações de comportamento, as quais são entidades físicas. A segunda premissa conduz à conclusão de que entidades não físicas não podem ser causa de qualquer entidade física. Visto que estados mentais são caracterizados como tipicamente causando entidades físicas, não podem ser não físicos. Assim, a conclusão do argumento causal de Lewis é a de que estados mentais sejam físicos.

É de algum interesse que Lewis ofereça versões resumidas do argumento em questão que não correspondem diretamente a ele. Em Lewis (1966, p. 17), lemos:

Meu argumento é esse: A característica definitiva de qualquer (espécie de) experiência como tal é seu papel causal, sua síndrome de causas e efeitos mais típicos. Mas nós materialistas acreditamos que esses papéis causais que pertencem às experiências por necessidade analítica pertençam de fato a certos estados físicos.

Em Lewis (1999 [1972] & 1999 [1994]), o esquema em favor de identificações entre um estado mental e o estado físico que executa seu papel funcional é apresentado como resumindo o argumento acima:

estado mental M = o ocupante do papel-m (por análise)  
 estado físico P = o ocupante do papel-m (pela ciência)  
 portanto, M = P (Lewis, 1999 [1994], p. 303 & 1999 [1972], p. 248s.)<sup>13</sup>.

As versões resumidas se assemelham, mas parecem oferecer uma caracterização pouco fiel do denominado argumento causal de Lewis. Com efeito, este não enfatiza que, em identidades psicofísicas, um estado mental seja idêntico a uma entidade física cuja natureza é ao menos parcial e indiretamente determinada. A partir das versões resumidas, reconhecemos que a entidade física em questão deva ser de uma espécie

---

<sup>13</sup> Seguindo Lewis, Jackson (2012) reconhece essa implicação como sendo o argumento de Lewis em favor do fisicismo de tipos.

que possa executar o papel funcional associado ao estado mental. Não creio que a diversidade nas formulações do(s) argumento(s) de Lewis ao fisicismo de tipos seja um ponto de grande relevância. Revela, no entanto, que a defesa de Lewis dessa forma de fisicismo é mais rica do que a descrição oferecida acima do argumento causal possa sugerir.

A identificação psicofísica envolve, segundo Lewis, passos *a priori* e *a posteriori*. De modo *a priori*, caracterizaríamos o papel funcional associado a um estado mental. Por meio de análise conceitual, podemos explicitar o conjunto de causas e efeitos típicos assim como as relações com outros estados mentais que a psicologia popular implicitamente atribui aos referentes de um termo para estado mental. Por meio de investigações *a posteriori*, por outro lado, podemos descobrir qual entidade não mental executa o papel funcional de um estado mental. Com base no argumento causal de Lewis, as entidades que executam tais papéis funcionais têm que ser físicas.

Como Place e Smart, Lewis concebe identificações psicofísicas como sendo contingentes. Caracteriza essa forma de contingência como a variação em mundos possíveis do tipo físico cujos *tokens* executam o papel funcional próprio de um estado mental. No entanto, sua defesa do fisicismo de tipos é mais sólida e convincente do que as razões oferecidas por Place ou Smart em favor da teoria da identidade, as quais envolvem tomar identidades psicofísicas como hipóteses científicas razoáveis ou propô-las a partir de princípios de parcimônia e simplicidade. O argumento causal de Lewis e os esquemas que, por meio de papéis funcionais, justificam identificações psicofísicas conferem maior plausibilidade ao fisicismo de tipos. No restante deste capítulo, argumentarei que as objeções de Putnam, Fodor, Davidson e Kripke descritas acima não ameaçam a plausibilidade do fisicismo de tipos tal como concebido por Lewis.

Há diferentes modos de conceber o argumento da realizabilidade múltipla, e nem todos são incompatíveis com a caracterização do fisicismo de tipos feita por Lewis. Caso consista na possibilidade de que *tokens* de

diferentes tipos executem o papel funcional de um estado mental em diferentes mundos possíveis, a realizabilidade múltipla é inofensiva. Com efeito, essa possibilidade é reconhecida por Lewis como uma razão em favor da tese da contingência de identidades psicofísicas. Alguns aspectos problemáticos, no entanto, parecem surgir se a realizabilidade múltipla for concebida como a possibilidade de que *tokens* de mais de um tipo físico possam ocupar, no mundo atual, o papel funcional associado a um mesmo tipo de estado mental. Essa formulação do argumento é ilustrada pela afirmação de Putnam de que dor em mamíferos, répteis e polvos possa ser realizada por processos nervosos de diferentes tipos.

Inicialmente, Lewis propunha que termos seriam sem denotação se os papéis funcionais que os definem fossem multiplamente realizados (Lewis, 1970, p. 434 & 1999 [1972], p. 252). A posição de que nenhuma entidade mereça o termo para um estado mental multiplamente realizado implica em um eliminativismo acerca do estado mental em questão. Desse modo, se entidades de diferentes tipos ocupassem a função de dor, então não haveria nada que fosse de fato dor. A solução de Lewis para essa estranha conclusão consiste em circunscrever tipos de estados mentais a populações ou espécies (Lewis, 1969, p. 25, 1983 [1980], p. 126 & 1999 [1994] p. 305s.). Como efeito, é muito pouco plausível que *tokens* de um mesmo tipo físico executassem o papel funcional de dor em diferentes espécies (Lewis, 1969, p. 25). Mas isso justifica apenas que devemos esperar que os teóricos da identidade reconheçam que estados cerebrais de diferentes tipos instanciem dor em diferentes espécies<sup>14</sup>. Desse modo, identidades psicofísicas devem ser concebidas como específicas a espécies (*species specific*). Lewis reformula o esquema em favor de identidades psicofísicas para tornar explícito o caráter específico a espécies de sua conclusão. Ao invés de “M=P”, a conclusão do esquema inferencial deve

---

<sup>14</sup> Lewis ilustra essa posição por meio do caso análogo dos números vencedores de loterias semanais (Lewis, 1969, p. 25). A identidade entre o 17 e o número vencedor desta semana não compromete a identidade entre o 28 e o número vencedor da semana passada.

ser “M-em-K=P” se investigações científicas revelarem que P ocupa o papel-M na espécie K (Lewis, 1999 [1994], p.305).

Lewis (1999 [1994]) abandona a posição de que os termos para estados mentais, ou os termos-T em geral, que sejam multiplamente realizados devam ser sem denotação. Propõe que se um termo para estado mental seja realizado por entidades de mais de um tipo, esse possa ser um caso em que o termo tenha uma referência ambígua (Lewis 1999, [1994], p. 301). Algum grau de disjunção parece ser aceitável em identidades entre tipos mentais e físicos. Entretanto, esse afrouxamento das exigências não permite identidades com tipos “extremamente disjuntivos” (*wildly disjunctive*), como Fodor caracteriza proferimentos de identidade entre tipos que pertençam a diferentes níveis da realidade (Fodor, 1980 [1974], p. 124). Em tais casos, um proponente do fisicismo de tipos pode ou bem procurar identidades específicas a espécies ou abraçar o eliminativismo.

A proposta de identidades específicas a espécies entre estados mentais e tipos físicos pode ser interpretada como acarretando a negação de que haja um aspecto geral comum que diga respeito aos estados mentais específicos a diferentes espécies. Para ilustrar esse ponto, consideremos novamente os casos de dor. Dado que dor em mamíferos seria uma coisa e dor em répteis, outra, podemos nos perguntar se não haveria nada em comum entre esses estados mentais. De acordo com Lewis, há algo que animais de diferentes espécies compartilham quando têm dor, a saber, a propriedade de estar com dor (*being in pain*) (Lewis, 1999 [1994], p. 307; Jackson, 2012, p. 161s.). A propriedade de estar com dor é distinta do estado de dor. A primeira é a propriedade de ter um estado que ocupe o papel funcional de dor. O estado de dor, por outro lado, é o estado que ocupa o mencionado papel funcional. Desse modo, apesar de dor ser diferente em mamíferos, répteis e polvos em decorrência de diferentes tipos físicos ocuparem o papel funcional de dor, esses animais podem compartilhar a propriedade de estar com dor, i.e., a propriedade de haver algo que execute neles tal função.

O holismo do mental e o papel que a racionalidade exerce sobre o conjunto de nossos estados mentais figuram como os principais elementos da crítica de Davidson à tese de que haja identidades entre tipos mentais e tipos físicos. Parece-me, no entanto, que as identidades propostas pelo fisicismo de tipos não o tornem, necessariamente, incompatível com as considerações sobre a interdependência de estados mentais e sua regulação pela racionalidade. Apesar de Lewis não ter enfrentado a objeção de Davidson explicitamente, creio que possamos encontrar as bases para uma resposta adequada em sua caracterização do fisicismo de tipos. De acordo com sua abordagem, podemos extrair da psicologia popular caracterizações de estados mentais que nos permitem descobrir e justificar identificações psicofísicas. Seguindo o procedimento que fornece, a partir do postulado da psicologia popular (TF), a sentença de Ramsey e a sentença de Carnap de TF, Lewis conclui que os termos mentais sejam definidos através da seguinte fórmula:

$$m_1, \dots, m_n = X_1, \dots, X_n \text{TF}[X_1, \dots, X_n]$$

Nela, encontramos definições para cada um dos termos mentais empregados pela psicologia popular, as quais identificam um estado mental como o ocupante de um dado papel funcional. Lewis ressalta que tais definições são obtidas como um pacote completo (*a package deal*) em que os termos mentais são interdefinidos (Lewis, 1966, p. 21 & 1999 [1994], p. 299). Devemos recordar que um papel funcional não consiste apenas nas causas e efeitos típicos de um estado mental, mas também inclui as relações que, segundo a psicologia popular, devem vigorar entre o estado mental em questão e outros estados mentais.

Segundo Davidson (1994, p. 231s.), o holismo do mental consiste na relação de interdependência entre estados mentais. Essa relação é ilustrada, e.g., na dependência que a crença de que haja cerveja na geladeira mantém com a crença de que ao menos uma geladeira e uma cerveja existam. Relações de (inter-)dependência entre estados mentais devem figurar na caracterização de seus papéis funcionais específicos.

Relações dessa espécie são asseridas por princípios da psicologia popular com que nos orientamos de modo tácito. Por análise, podemos torná-las explícitas. Ao atribuir a alguém a crença de haja cerveja na geladeira, por exemplo, esperamos que tenha a crença de que ao menos uma cerveja e uma geladeira existam. Tal expectativa dificilmente é consciente, mas nos surpreenderíamos enormemente caso se revelasse falsa. Visto que os termos mentais seriam, segundo Lewis, interdefinidos a partir de seus papéis funcionais, em que as relações entre estados mentais apontadas pela psicologia popular são reconhecidas, concluo que as identidades entre estados mentais e os ocupantes físicos de seus papéis funcionais sejam compatíveis com a tese do holismo do mental.

O papel regulatório que Davidson atribui à racionalidade na determinação de ocorrências de estados mentais se revela, do mesmo modo, compatível com a caracterização lewisiana do fisicismo de tipos. Lewis afirma que a psicologia popular “nos credita com um mínimo de racionalidade em nosso agir, crer e desejar”, o qual atua no estabelecimento de delimitações sobre quais estados mentais possamos instanciar nas mais variadas circunstâncias (1999 [1994], p. 320). Dado que as definições funcionais de termos mentais são extraídas do postulado da psicologia popular, elas terminam por incluir especificações sobre as formas de atuação da racionalidade para os diversos estados mentais nas mais diversas circunstâncias. Por empregar definições dessa espécie para justificar identidades psicofísicas, o fisicismo de tipos tal como caracterizado por Lewis preserva o papel regulador da racionalidade. Desse modo, revela-se intocado pela objeção de Davidson mencionada acima.

O funcionalismo analítico também atua de forma significativa na resposta de Lewis ao argumento de Kripke contra o fisicismo de tipos. Como vimos, a referida forma de funcionalismo identifica um estado mental com aquilo que ocupa (ou executa) seu papel funcional. Assim, a execução de um tal papel por algo físico implica em uma identidade psicofísica. Como Place e Smart, Lewis compreende identidades

psicofísicas como contingentes. Ele caracteriza essa contingência como a possibilidade de entidades que pertençam a diferentes tipos físicos executarem em diferentes mundos possíveis (ou até mesmo no mundo atual, em casos de referência ambígua do termo mental) o papel funcional associado a um estado mental.

A tese da contingência de identidades psicofísicas se contrapõe à primeira premissa do argumento de Kripke contra o fisicismo de tipos. Como elucidada acima, a primeira premissa afirma que, em virtude de proferimentos de identidades psicofísicas empregarem designadores rígidos, tais proferimentos devem ser necessariamente verdadeiros caso sejam verdadeiros. A segunda premissa afirma que temos uma razão para assumir que proferimentos de identidades psicofísicas não sejam necessariamente verdadeiros, a saber, a aparência de sua contingência. Como conclusão, o argumento diz que temos uma razão para assumir proferimentos de identidades psicofísicas e o fisicismo de tipos como falsos.

Lewis evita a conclusão do argumento de Kripke ao negar um aspecto de sua primeira premissa. Nega que termos para estados mentais, mesmo aqueles que se referem a estados de consciência, como “dor”, “ver vermelho” etc., sejam designadores rígidos. Lewis reconhece que diverge de Kripke em suas intuições e oferece um argumento em favor de sua posição (Lewis, 1999 [1994]. p. 304), o qual elucidado abaixo. A partir da caracterização de dor como o que ocupa o papel de dor, Lewis nos pede para considerar uma ocasião em que sentimos dor. Nessa ocasião, algo ocupou o papel funcional de dor e, dessa forma, mereceu a designação “dor”. Devemos, em seguida, considerar uma situação hipotética em que algo distinto, outro estado, ocupe o papel de dor. Apesar da diferença no ocupante do papel de dor, a situação hipotética deve se assemelhar em todos os aspectos possíveis com a situação que vivemos no mundo atual, de forma que não conseguiríamos distinguir facilmente as duas situações. No entanto, Lewis nos diz, a tese da designação rígida de “dor” teria como consequência que as duas situações fossem facilmente distinguidas, uma

delas envolveria dor e a outra, não. Por se tratar de uma consequência absurda, Lewis rejeita que “dor” e outros termos para estados mentais sejam designadores rígidos. O argumento de Lewis ganha força por não especificar que os ocupantes do papel de dor nas duas situações sejam estados físicos. Revela-se válido também se os diferentes estados que ocupam o papel de dor na situação do mundo atual e na situação hipotética forem não físicos, como propõem os críticos do fisicismo, entre eles, Kripke. Se “dor” for um designador rígido, o estado não físico que ocupa o papel de dor na situação hipotética descrita acima não será dor, uma vez que é um estado distinto do estado não físico que designamos com o termo “dor” na situação do mundo atual.

A posição de que termos para estados mentais sejam designadores não rígidos não implica na negação do caráter fenomenal de estados de consciência. Para Lewis, é incontroverso que “ter uma dor e sentir uma dor seja uma e a mesma coisa” (Lewis, 1983 [1980], p. 130). Entretanto, a não rigidez de “dor” implica contingência entre o caráter fenomenal da dor com relação àquilo que é dor em razão de ser o ocupante do papel funcional de dor. Em outras palavras, “[d]or poderia não ter sido dor” (Lewis, 1983 [1980], p. 125). Lewis elucida essa afirmação aparentemente contraditória ao ressaltar que o estado físico que executa o papel funcional de dor, poderia não o fazer. Outros estados físicos poderiam executá-lo, e, assim, “[a]lgo que não é dor poderia ter sido dor” (Lewis, 1983 [1980], p. 125). Fred Feldman (1974) defende que a sensação de dor possa ser uma propriedade contingente da dor, no sentido de que algo poderia ser dor em um mundo possível, mas não envolver a sensação de dor. Deve ficar claro que a espécie de contingência proposta por Lewis é de natureza distinta. De acordo com ele, é o estado físico cujas ocorrências executam o papel funcional de dor que está relacionado de forma contingente com a sensação de dor. Se, em um mundo possível, as ocorrências de um tal estado físico não envolvam sensações de dor, então tais ocorrências não são ocorrências de dor, e o estado não é idêntico ao estado dor.

## Conclusão

Após apresentar o fisicismo de tipos como uma generalização da teoria da identidade de Place e Smart, elucidei as objeções de formulação que lhe foram levantadas por Putnam, Fodor, Davidson e Kripke. Tais objeções visavam revelar a falsidade do fisicismo de tipos ao demonstrar sua implausibilidade em decorrência do modo como foi formulado. Em seguida, apresentei a forma como Lewis caracterizou e argumentou em favor do fisicismo de tipos, enfatizando sua tese do funcionalismo analítico do mental. Por fim, explicitiei como Lewis respondeu ou poderia responder às objeções de formulação dos oponentes do fisicismo de tipos, ressaltando a grande relevância atribuída ao funcionalismo analítico na construção dessas respostas. Se quisermos representar o confronto de posições entre Lewis, de um lado, e Putnam, Fodor, Davidson e Kripke, de outro, como uma disputa esportiva, parece-me que o funcionalismo analítico possa ser descrito como a técnica que garantiu ao primeiro a vitória.

## Referências

- CARNAP, R. *Psychologie in Physikalischer Sprache*. **Erkenntnis**, Vol. 3, p. 107-42, 1932.
- DAVIDSON, D. *Mental Events*. In: BLOCK, N. (Org.). **Readings in the philosophy of psychology**. Vol.1. London: Methuen and Co.,1980 [1970], p. 107-19.
- \_\_\_\_\_. Davidson, Donald. In: GUTTENPLAN, S. (Org.). **A Companion to the Philosophy of Mind** (Blackwell Companions to Philosophy). Oxford: Basil Blackwell, 1994, p. 231-6.
- FODOR, J. A. *Special Sciences (or the Disunity of Science as a Working Hypothesis)*. In: BLOCK, N. (Org.). **Readings in the philosophy of psychology**. Vol.1. London: Methuen and Co., 1980 [1974], p. 120-33.
- FELDMAN, F. *Kripke on the Identity Theory*. **The Journal of Philosophy**, Vol. 71, No. 18, Seventy-First Meeting of the American Philosophical Association Eastern Division (Oct. 24, 1974), p. 665-76, 1974.

GOUVEA, R. A. S. Physical objects, physical theories and what is physical in physicalism.

**Dissertatio**, Vol. 44, p. 39-64, 2016.

\_\_\_\_\_. Physicalism without identity. **Trans/Form/Ação**, Vol. 43, No. 2, p. 253-80, abr./jun., 2020.

KRIPKE, S. **Naming and necessity**. Oxford: Blackwell Publishing, 1981 [1972].

\_\_\_\_\_. Identity and Necessity In: KIM, J., KORMAN, D. Z. & SOSA, E. (Org.). **Metaphysics: an anthology**. Wiley-Blackwell, 2012 [1971], p. 122-40.

JACKSON, F. In defence of the identity theory mark *I*. In: GOZZANO, S. & HILL, C. (Org.).

**New perspectives on type identity: The mental and the physical**. Cambridge & New York: Cambridge University Press, 2012. p. 150-66.

LEWIS, D. An Argument for the Identity Theory. **The Journal of Philosophy**, Vol. 63, No.

1, p. 17-25, jan., 1966.

\_\_\_\_\_. Review of Art, Mind, and Religion. **The Journal of Philosophy**, Vol. 66, No. 1, p. 22-7, jan., 1969.

\_\_\_\_\_. How to Define Theoretical Terms. **The Journal of Philosophy**, Vol. 67, No. 13, p. 427-46, jul., 1970.

\_\_\_\_\_. Mad Pain and Martian Pain. In: \_\_\_\_\_. **Philosophical papers**. Vol. 1. New York & Oxford: Oxford University Press, 1983 [1980], p. 122-30.

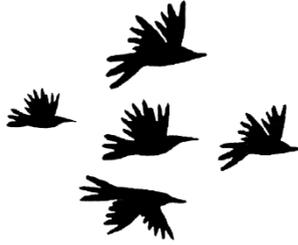
\_\_\_\_\_. Psychophysical and theoretical identifications. In: \_\_\_\_\_. **Papers in metaphysics and epistemology** (Cambridge Studies in Philosophy). Cambridge: Cambridge University Press, 1999 [1972], p. 248-61.

\_\_\_\_\_. Reduction of mind. In: \_\_\_\_\_. **Papers in metaphysics and epistemology** (Cambridge Studies in Philosophy). Cambridge: Cambridge University Press, 1999 [1994], p. 291-324.

NEURATH, O. **Empirische Soziologie: Der wissenschaftliche Gehalt der Geschichte und Nationalökonomie**. Vienna: Verlag von Julius Springer, 1931a.

\_\_\_\_\_. Physicalism: The Philosophy of the Viennese Circle. **The Monist**, Vol. 4, p. 618- 23, out., 1931b.

- \_\_\_\_\_. Physikalismus. In: HALLER, R. & RUTTE, H. (Org.) **Otto Neurath: Gesammelte philosophische und methodologische Schriften Band 1**. Vienna: Verlag Hölder-Pichler-Tempsky, 1981 [1931], p. 417-21.
- PLACE, U. T. Is Consciousness a Brain Processes? **British Journal of Psychology**, Vol. 47, p. 44-50, 1956.
- PUTNAM, H. The Nature of Mental States. In: \_\_\_\_\_. **Mind, Language and Reality, Philosophical Papers**. Vol. 2. Cambridge: Cambridge University Press, 1975 [1967], p. 429-40.
- SMART, J. J. C. Sensations and Brain Processes. **The Philosophical Review**, Vol. 68, No. 2, p. 141-56, apr., 1959.
- WETZEL, L. Types and Tokens. In: ZALTA, E. N. (Org.) **The Stanford Encyclopedia of Philosophy** (Spring Edition 2011). Disponível em: <<http://plato.stanford.edu/archives/spr2011/entries/types-tokens/>> Acesso em: 01 julho 2020.



## Quine e a dispensabilidade da identidade

*Hudson dos Anjos Benevides*<sup>1</sup>

### Introdução

A identidade é um daqueles tópicos que atravessa campos dentro da filosofia. Os debates envolvendo essa noção, aparentemente simples, tomam diferentes formas. Por exemplo, em metafísica há uma longa tradição pondo em questão se podemos ou não tomar banho no mesmo rio duas vezes. Questões de *identidade pessoal*, *critérios de individuação* e *identidade entre mundos possíveis* estão entre os principais interesses de filósofos que dirigem seus esforços na tentativa de especificar quais são as categorias mais básicas da realidade. O que nos ocupará, daqui em diante, é a questão se a identidade é indispensável para o vocabulário lógico. Em particular, iremos analisar os argumentos que Quine (1986) apresenta a favor da dispensabilidade da identidade. Como veremos, com os devidos esclarecimentos, a tese de que a identidade pode ser dispensável, nos moldes em que Quine propõe, precisa ser revisada.

---

<sup>1</sup> Doutorando em Filosofia pelo Programa de Pós-graduação Lógica e Metafísica da Universidade Federal do Rio de Janeiro (UFRJ). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) durante mestrado realizado entre 2016-2017 na Universidade Federal do Rio Grande do Norte (UFRN). E-mail: [benevideshudson@gmail.com](mailto:benevideshudson@gmail.com)

A questão do caráter lógico da identidade é relevante para a discussão mais ampla sobre os limites da lógica. Diversos autores têm tentado lidar com esse problema, mas um critério amplamente satisfatório para distinguir o que é uma noção lógica do que não é ainda se mostra um empreendimento em aberto. Alguns autores, que buscam uma abordagem modelo-teórica, optam por invariância sob permutação (TARSKI, 1986; BONNAY, 2006). Outros, alinhados a teoria das demonstrações, usam as regras de inferências como elemento característico das constantes lógicas (GENTZEN, 1969; SCHROEDER-HEISTER, 1984). Há ainda aqueles que, céticos quanto à existência de um critério criado por meio dessas metodologias, propõem uma demarcação pragmática, baseada nos propósitos que os lógicos podem ter ao criar uma lógica (WARMBROD, 1995; GÓMEZ-TORRENTE, 2002). No entanto, como Gómez-Torrente (2002) observa, poucos autores têm buscado uma análise conceitual das noções lógicas caso a caso. O presente texto busca preencher parte dessa lacuna ao analisar o caso da identidade na lógica.

O presente texto está organizado da seguinte forma. Começamos esclarecendo o que será tomado como o sentido lógico de identidade. Em seguida, será apresentada algumas formas de defender o caráter indispensável, em especial, aqueles que Dummett (1973) aponta. Na sequência, a posição de Quine (1986) sobre a identidade é esclarecida, para que em seguida se apresente as dificuldades que essa abordagem enfrenta.

### **O sentido lógico de “identidade”**

De um ponto de vista intuitivo, dizer que um objeto “a” é idêntico a “b” é dizer que “a” e “b” são o mesmo, o que pode parecer um truísmo. Contudo, a noção de “mesmidade” é ambígua; pode ser compreendida ao menos de dois modos. Dizemos que certos objetos são *qualitativamente idênticos* se eles compartilham apenas algumas propriedades. Quando há coincidência entre todas as propriedades dizemos que a identidade é *identidade numérica*. Um bom exemplo são gêmeos idênticos. Dizemos

que eles são idênticos porque compartilham um grande número de propriedades, mas ainda assim é certo concebê-los como indivíduos diferentes. Em nossa terminologia, gêmeos idênticos são apenas qualitativamente idênticos. Por outro lado, é inegável que cada um deles partilha exatamente as mesmas propriedades consigo mesmo. Chamamos de identidade numérica a *relação que uma coisa tem consigo mesmo e com nada mais*. É esse o sentido de identidade que empregamos ao longo do texto.

Em lógica de primeira ordem, os seguintes axiomas são incorporados ao sistema quando a identidade é adicionada ao conjunto de constantes lógicas:

$$(\alpha_1): \forall x(x = x)$$

$$(\alpha_2): \forall x \forall y (x = y \rightarrow (Fx \leftrightarrow Fy))$$

O axioma  $(\alpha_1)$  expressa a reflexividade da identidade que, por sua vez, captura a ideia de que *tudo é idêntico a si mesmo*.  $(\alpha_2)$  é o princípio leibniziano de *indiscernibilidade dos idênticos* de que se “ $x$ ” é idêntico a “ $y$ ”, então o que é verdade acerca de “ $x$ ” também é verdade sobre “ $y$ ”. Como é evidente, a identidade está associada ao símbolo “=”, que advém da matemática. Não apenas o símbolo, mas algumas propriedades que são familiares aos matemáticos também são atribuídas a identidade como, por exemplo, reflexividade, simetria, transitividade.

Em um sistema de dedução natural para a lógica clássica de primeira ordem – doravante  $LPO^=$  – as regras de inferência para a identidade são as seguintes:

$$(=I) \frac{}{t = t} \quad (=E) \frac{x = y \quad Fx}{Fy}$$

A regra de introdução da identidade  $(=I)$  não depende de suposições. Em  $LPO^=$  essa regra pode ser introduzida em qualquer passo de uma dedução. A regra de eliminação da identidade  $(=E)$  é a expressão da

indiscernibilidade dos idênticos mencionada acima, mas como uma regra de substituição. Se em uma dedução é possível derivar uma identidade e, em outro contexto, obtém-se uma sentença na qual ocorre um dos termos da identidade previamente estabelecida, então é justificável a substituição pelo outro termo da identidade. Como mera ilustração do que foi descrito, suponha que se afirme que “Lewis Carroll é o autor de *Alice no País das Maravilhas*” e que, posteriormente, estabeleça-se que “Lewis Carroll é Charles Dodgson”; por ( $=_E$ ) estamos autorizados a concluir que “Charles Dodgson é o autor de *Alice no País das Maravilhas*”.

Além disso, identidade também é usada com o propósito de se regimentar algumas expressões da linguagem natural, tais como, ‘pelo menos/no máximo  $n$  indivíduos tais que...’, ‘o único  $x$  tal que...’. As sentenças,

- (1) Pelo menos três pessoas leram o texto,
- (2) Maria comprou no máximo dois livros,
- (3) O autor de *A revolução dos bichos* é jornalista,

são formalizadas em  $LPO$ , respectivamente, como

$$(1^*) \exists x \exists y \exists z (\sim x = y \wedge \sim x = z \wedge \sim y = z \wedge \exists u (R_x u \wedge R_y u \wedge R_z u))$$

$$(2^*) \exists x \exists y (C_m x \wedge C_m y \wedge \forall z (C_m z \rightarrow (x = z \vee y = z)))$$

$$(3^*) \exists x (A_x \wedge \forall y (A_y \rightarrow x = y) \wedge J_x)$$

É importante notar que, com relação a (3), (3<sup>\*</sup>) é apenas uma possibilidade de regimentação, aquela que deriva do tratamento russelliano das descrições definidas. As expressões mencionadas acima não podem ser apresentadas em lógica de primeira ordem sem identidade. Desse modo,  $LPO^=$  possui maior poder expressivo do que  $LPO$ . Aparentemente, se estamos interessados na cardinalidade do domínio do discurso, então é preciso levar a identidade em consideração. Entretanto,

seriam as considerações feitas, até então, suficientes para tratar “=” como uma constante lógica? Analisaremos a seguir um argumento em favor da logicidade da identidade.

### Frege–Dummett: a indispensabilidade da identidade<sup>2</sup>

De acordo com Dummett (1973, p. 542), foi Frege o primeiro a tratar a identidade como uma noção lógica e dentre as principais motivações está o fato de que a identidade é pressuposta em qualquer definição aceitável e, desse modo, a identidade seria indispensável. Como Dummett observa,

[...] de acordo com Frege, identidade é indefinível visto que ela é necessária para a formulação de qualquer definição: uma definição, ao menos uma definição explícita, deve sempre assumir a forma de uma estipulação da verdade de um enunciado de identidade; e, em seus escritos após a *Grundlagen*, Frege não reconheceu a legitimidade de nenhum outro tipo de definição além de uma definição explícita. É verdade que, na *Grundlagen*, Frege fala da lei de Leibniz como uma ‘definição’ da identidade, e diz que ele irá adotá-la como sua: contudo, nos dois sistemas formais, aquele da *Begriffsschripte* aquele da *Grundgesetze*, a identidade é um primitivo, e a tese de que a identidade é indefinível é o fundamento que ele fornece, em *FunctionundBegriffe* em outro lugar, para considerá-la desse modo(DUMMETT, 1973, p. 542-3)<sup>3</sup>.

De fato, se qualquer definição envolve a identidade, tentar defini-la seria já uma petição de princípio. Não à toa, Frege toma a identidade como uma noção primitiva. Contudo, Dummett argumenta que a tese fregeana da indefinibilidade da identidade não é convincente. Segundo Dummett (1973, pg. 543), Frege usava a identidade para cumprir também o papel do

<sup>2</sup> Para a conveniência do leitor, as citações foram traduzidas ao longo do texto, mantendo-se os trechos originais em notas de rodapé.

<sup>3</sup> “[...] according to Frege, identity is undefinable, since it is required for the formulation of any definition: a definition, at least an explicit definition, must always take the form of a stipulation of the truth of an identity-statement; and, in his writings after *Grundlagen*, Frege recognized the legitimacy of no other kind of definition than an explicit one. It is true that, in *Grundlagen*, Frege speaks of Leibniz’s law as a ‘definition’ of identity, and says that he will adopt it as his own: in both formal systems, however, that of *Begriffsschrift* and that of *Grundgesetze*, identity is primitive, and the thesis that identity is undefinable is the ground he gives, in *Function und Begriff* and elsewhere, for taking it so.”

bicondicional. Não iremos levar essa discussão adiante, mas é importante notar que a tese de indefinibilidade é recorrente na literatura – veja, por exemplo, SAVELLOS (1990); MCGINN (2000); BUENO (2014). O problema que iremos tratar é o seguinte. Comparada às outras constantes lógicas, a identidade possui um comportamento peculiar. Por um lado, temos expressões, como constantes individuais e predicados, que são usadas de modo a produzir sentenças atômicas e que não fazem parte do vocabulário lógico. Conectivos sentenciais e quantificadores, por seu lado, são operadores usados para formar sentenças complexas a partir de sentenças atômicas. Como pode ser observado com os exemplos anteriormente descritos, o símbolo “=” combina duas constantes individuais para formar uma sentença atômica, exercendo o mesmo comportamento das expressões que compõem o vocabulário não-lógico. Em outras palavras, dentro desta divisão sintática, a identidade mantém-se aquilo que intuitivamente é: uma relação. Ora, se conectivos sentenciais e quantificadores pertencem a categoria de itens formadores de sentenças complexas e a identidade não, e tendo em mente o fato de que os primeiros são considerados constantes lógicas, então por que haveríamos de conceber a identidade como lógica, dado que sua categoria é distinta daqueles?

A pergunta que enunciamos acima só é um problema caso estejamos comprometidos com um *critério gramatical de logicidade*. De acordo com um critério gramatical, contam como lógicas as expressões que se comportam como operações formadoras de sentenças complexas. Embora tome o critério gramatical apenas como uma sugestão, Dummett (1973, p. 22) afirma que este é um princípio simples e preciso que distingue as constantes lógicas de outras expressões e que se adéqua à prática corrente, porque admite expressões como “necessariamente”, “possivelmente”, “deve”, i.e., expressões que quando adicionadas a linguagens formais deverão construir sentenças complexas a partir das sentenças atômicas<sup>4</sup>.

---

<sup>4</sup> O termo “gramatical”, portanto, é usado no sentido de gramática de linguagens formais e não como gramática da linguagem natural. Ressaltamos que a inclusão de Dummett como representante dessa vertente é devido a

Desse ponto de vista, é com essas expressões desse tipo que a lógica deveria estar comprometida. Com as outras expressões, sejam elas constantes individuais, predicados ou quaisquer outras, a lógica interessasse apenas esquematicamente:

[...] isto é, ela é comprometida com as regras gerais governando a subdivisão das palavras do primeiro tipo em diferentes categorias lógicas – nomes próprios, predicados unários, expressões relacionais, e assim por diante – e com os modos nos quais termos destes vários tipos diferentes podem ser postos juntos para formar sentenças atômicas; mas não com os sentidos de qualquer termo particular ou expressões destas diferentes categorias (DUMMETT, 1973, p. 22)<sup>5</sup>.

Ciente da dificuldade que deveria enfrentar com relação a “=” se insistir no critério gramatical, Dummett faz a seguinte concessão:

Chamemos de *condição de segundo nível* qualquer condição que, para algum domínio de objetos, é definida como sendo satisfeita ou o contrário por todo predicado que é, por sua vez, definido sobre este domínio de objetos. Dentre tais condições de segundo nível, podemos chamar de *condição quantificacional* (*quantifier condition*) qualquer uma que seja invariante sob cada permutação do domínio de objetos: i.e., para qualquer predicado ‘ $F(\xi)$ ’ e qualquer permutação  $\phi$ , ela satisfaz ‘ $F(\xi)$ ’ apenas no caso em que ela satisfaz o predicado que aplica-se a apenas aqueles objetos  $\phi(a)$ , onde ‘ $F(\xi)$ ’ é verdadeiro de  $a$ . Assim, nós permitimos também como sendo uma constante lógica qualquer expressão que, com a ajuda dos quantificadores universal e existencial e os operadores sentenciais, nos permite expressar uma condição quantificacional que não poderia ser expressa por meio daqueles dois quantificadores e dos operadores sentenciais sozinhos. Desse modo, o sinal da identidade é reconhecido, sob este critério como uma constante lógica, visto que ela nos

---

(MACFARLANE, 2015). Dummett não menciona o termo gramatical, mas apenas fala da construção de sentenças complexas passo-a-passo (*step-by-step*).

<sup>5</sup> “[...] that is, it is concerned with the general rules governing the subdivision of words of the first type into different logical categories—proper names, one-place predicates, relational expressions, and so forth—and with the ways in which words of these various different types can be put together to form atomic sentences; but not with the senses of any particular words or expressions of these different categories”.

permite expressar a condição de que um predicado se aplica a no máximo um objeto, que não poderia ser expressa sem ela. (DUMMETT, 1973, p. 22)<sup>6</sup>.

Alguns esclarecimentos para explicar a proposta de Dummett. Primeiramente, notamos que a letra grega ( $\xi$ ) é usada acima como uma variável, indicando o lugar de um argumento em um predicado. Uma *permutação* do domínio de objetos é qualquer *bijeção* do domínio em si mesmo. Lembramos que uma *bijeção* é uma função que é tanto *injetiva* (i.e., não existem objetos distintos no domínio da função que possuam a mesma imagem) quanto *sobrejetiva* (i.e., todos os elementos do contradomínio da função são imagem de algum elemento do domínio). *Grosso modo*, um conjunto é *invariante sob permutação* quando sua permuta devolve como resultado o próprio conjunto. Em termos de funções, isso quer dizer que o domínio e a imagem da função coincidem. Agora, suponha que quiséssemos propor uma condição ‘*existe no máximo n coisas...*’ (abreviada como  $\exists_{\leq n}$ ) que será satisfeita quando o número de objetos que satisfazem um determinado predicado  $Fx$  em um domínio  $D$  não ultrapassa o número  $n$ . Seguindo o raciocínio de Dummett, essa só poderá ser uma condição quantificacional se for verdadeira em toda interpretação induzida pela permutação do domínio. Ou seja, se  $D^*$  é uma permuta de  $D$ ,  $\exists_{\leq n} xFx$  tem de ser verdadeira em  $D^*$ , onde a nova interpretação de  $Fx$  consiste exatamente das imagens dos objetos de  $D$  obtidos por meio da permutação.

A justificativa de Dummett para a logicidade da identidade pode ser resumida como: (i) a condição ‘*no máximo um objeto tal que...*’ é invariante sob permutação, mas caso não seja tomada como noção primitiva, (ii) essa condição só pode ser formulada com a ajuda da identidade, então (iii) a

---

<sup>64</sup>“Let us call a second-level condition any condition which, for some domain of objects, is defined, as being satisfied or otherwise, by every predicate which is in turn defined over that domain of objects. Among such second level conditions, we may call a quantifier condition any which is invariant under each permutation of the domain of objects: i.e. for any predicate ‘ $F(\xi)$ ’ and any permutation  $\phi$ , it satisfies ‘ $F(\xi)$ ’ just in case it satisfies that predicate which applies to just those objects  $\phi(a)$ , where ‘ $F(\xi)$ ’ is true of  $a$ . Then we allow as also being a logical constant any expression which, with the help of the universal and existential quantifiers and the sentential operators, allows us to express a quantifier condition which could not be expressed by means of those two quantifiers and the sentential operators alone. Thus, the sign of identity is recognized, on this criterion, as a logical constant, since it allows us to express the condition that a predicate applies to at most one object, which cannot be expressed without it.”

identidade deve contar como constante lógica. O curioso é que invariância sob permutação foi o critério de logicidade proposto por Tarski, que, por si só, conta a identidade como uma noção lógica, assim não precisaríamos de uma concessão. Dado que os objetivos de Dummett (1973) são outros que não o de fornecer uma teoria de logicidade para as expressões que a lógica deveria lidar, é difícil estabelecer o que exatamente o critério gramatical pretende captar; se uma noção de tópico-neutralidade ou apenas a ubiquidade de certos termos linguísticos. De todo modo, ao menos é claro que, assim como Frege, Dummett considerou, por uma perspectiva diversa, o mesmo fato: a indispensabilidade da identidade. Entretanto, quão forte é a tese da indispensabilidade da identidade? Será a indispensabilidade suficiente para favorecer o caráter lógico da identidade? Adiante analisamos a tese de Quine de que a identidade é dispensável.

### **Quine e a logicidade da identidade**

De acordo com Quine, muito embora existam razões para tratar a identidade como uma constante lógica, de um ponto de vista lógico podemos propor um substituto que satisfatoriamente cumpra o seu papel. Se isso for correto, então estamos a um passo de viver sem a identidade, o que sugere que a inclusão da identidade na lógica, até então, não passa de acidente histórico. Consequentemente, critérios de logicidade que contam a identidade como noção lógica deveriam ser revisados. As afirmações que servem de fundamento para a tese quineana da dispensabilidade da identidade são apresentadas adiante.

Quine é outro grande representante do critério gramatical para logicidade. Em sua abordagem, Quine (1986, p. 26-7) associa os termos lógicos de uma linguagem lógica às suas *partículas* gramaticais, que são os itens linguísticos responsáveis pela construção de sentenças complexas a partir de sentenças mais simples. São essas partículas que constituem a estrutura lógica das sentenças, elemento central em sua formulação da

noção de verdade lógica. Os outros itens da linguagem fazem parte do *léxico*, que abarca, por exemplo, letras esquemáticas. Embora tenha endossado o critério gramatical como justificativa para a logicidade, Quine (1986) escolheu uma abordagem diferente daquela de Dummett ao abordar a identidade.

Primeiramente, abordaremos por que Quine se preocupa com a identidade, e nosso ponto de partida é a noção de verdade lógica. Uma das propostas de Quine é considerar que uma verdade lógica nada mais é que “uma sentença cuja verdade é assegurada por sua estrutura lógica” (QUINE, 1986, pg. 49). A estrutura lógica depende da disposição dos operadores lógicos e dos itens que pertencem ao léxico, como predicados e variáveis. A sentença ‘ $\sim \exists x(\text{Filosofo}(x) \wedge \sim \text{Filosofo}(x))$ ’ é uma instância de uma verdade lógica conhecida como *princípio da não-contradição*, em primeira ordem formalizado como ‘ $\sim \exists x(Fx \wedge \sim Fx)$ ’. Quine (1986, p. 49) então propõe que verdade lógica pode ser entendida em termos de substituição, a saber, que “uma sentença é logicamente verdadeira se ela permanece verdadeira sob todas as mudanças de seus predicados”. De um ponto de vista linguístico, a identidade (=) é um predicado binário. Assim, verdades da teoria da identidade como, por exemplo, “ $x = x$ ”, não seriam verdades lógicas, uma vez que poderíamos substituir (=) por outros predicados que falsificam a afirmação – basta substituir o símbolo de identidade (=) por (>) da relação “ser maior que” e ver que isso é o caso. Então, a preocupação de Quine é saber se deveríamos encarar a (=) como uma noção pertencente à lógica ou se deveríamos considerá-la noção extra-lógica, pertencente, por exemplo, à matemática. É à luz dessas considerações que Quine elabora os argumentos expostos abaixo.

*A teoria da identidade é completa.* Como fato bem conhecido, a lógica clássica de primeira ordem possui a (meta-)propriedade de completude, o que significa que se um argumento é semanticamente válido, então há uma demonstração formal que testemunha esse fato. Quando os axiomas da identidade ( $\alpha_1$ ) e ( $\alpha_2$ ) são adicionados à lógica clássica, essa propriedade ainda se mantém válida. Quine usa isso com o objetivo de contrastar a

teoria da identidade com a teoria dos números. A teoria dos números não possui um formalismo dedutivo completo como foi demonstrado pelo famoso teorema da incompletude proposto por Gödel. Do ponto de vista quineano, esse é um aspecto no qual “a teoria da identidade parece um vizinho próximo da lógica” (QUINE, 1986, p. 62).

*Identidade é universal.* Sobre isso Quine afirma o seguinte:

Outro aspecto no qual a teoria da identidade parece mais próxima à lógica do que à matemática é a universalidade: ela trata todos os objetos imparcialmente. Qualquer teoria na verdade pode do mesmo modo ser formulada com variáveis gerais alcançando tudo, mas ainda assim os únicos valores de variáveis que importam para a teoria dos números, por exemplo, ou teoria dos conjuntos, são os números e os conjuntos; enquanto que a teoria da identidade não conhece preferências (QUINE, 1986, p.62)<sup>7</sup>.

Além da universalidade da identidade, entendida aqui como aquilo que ela aplica-se a tudo, como sugerido na citação acima, essa passagem também introduz uma noção de tópico-neutralidade – além de se aplicar a absolutamente tudo, a identidade também é tópico neutra, no sentido de ser indiferente aos objetos do discurso<sup>8</sup>. Como MacFarlane (2015) observa, mesmo considerando que a aritmética é sobre números, Frege, ainda assim, considerou-a com aplicabilidade geral.

*Identidade é unicamente caracterizável.* Intuitivamente, a ideia de que uma constante lógica é unicamente caracterizável significa que se um conjunto de axiomas ou regras é fornecido para uma constante  $\#_1$ , então uma versão duplicada desses mesmos axiomas ou regras que caracterize uma constante  $\#_2$  deverá resultar no fato de que as fórmulas construídas com  $\#_1$  e  $\#_2$  a partir dos mesmos componentes deverão ser equivalentes. Em particular, Quine (1986, p. 62-3) observa que isso acontece com a identidade. Definir duas noções de identidade, digamos “ $=_1$ ” e “ $=_2$ ”,

---

<sup>7</sup> “Another respect in which identity theory seems more like to logic than mathematics is universality: it treats of all objects impartially. Any theory can indeed likewise be formulated with general variables, ranging over everything, but still the only values of the variables that matter to number theory, for instance, or set theory, are the numbers and the sets; whereas identity theory knows no preferences.”

<sup>8</sup> Essa noção de tópico-neutralidade é proposta por MacFarlane (2015).

satisfazendo os mesmos axiomas  $(\alpha_1)$  e  $(\alpha_2)$ , nos conduz ao fato de que “ $=_1$ ” e “ $=_2$ ” irão coincidir em suas ocorrências. É fácil ver que isso é o caso, basta que se demonstre que “ $x =_1 y$ ” implica que “ $x =_2 y$ ” e vice-versa. Utilizaremos  $(\alpha_1^=1)$  e  $(\alpha_2^=1)$  para se referir aos axiomas da identidade “ $=_1$ ” e os respectivos para a “ $=_2$ ” com a devida indicação. Assim sendo, suponha que (i)  $x =_1 y$ . Como  $(\alpha_2^=1)$  é válido para qualquer predicado  $F$ , temos que (ii)  $x =_1 y \rightarrow (x =_2 x \leftrightarrow x =_2 y)$ . De (i) e (ii), por *modus ponens*, obtemos (iii)  $(x =_2 x \leftrightarrow x =_2 y)$ . Do axioma  $(\alpha_1^=2)$  e (iii), pelo significado da bi-implicação, segue-se que (iv)  $x =_2 y$ , que é parte da conclusão que pretendemos alcançar<sup>9</sup>. Para a conclusão de que  $x =_2 y$  implica que  $x =_1 y$  basta que se modifique a demonstração trocando as ocorrências de  $=_1$  por  $=_2$ , o mesmo se aplicando a seus respectivos axiomas. Portanto, parece razoável a tese de que a identidade é unicamente caracterizável.

*Identidade é redutível ou dispensável.* Quine argumenta que não precisamos de suposições extras para obter a identidade, no sentido de que o conjunto de axiomas ou mesmo os argumentos anteriormente esboçados são totalmente dispensáveis. A proposta aqui é a de produzir um substituto para a identidade. A estratégia usada por Quine é reduzir a noção de identidade à noção de indistinguibilidade a partir da combinação exaustiva dos predicados da linguagem que temos em mãos. Como exemplo, Quine considera uma linguagem com apenas quatro predicados, sendo A um predicado unário, B e C predicados binários e D um predicado ternário. A combinação mencionada segue os seguintes moldes:

$$(Q)Ax \leftrightarrow Ay \wedge \forall z(Bzx \leftrightarrow Bzy \wedge Bxz \leftrightarrow Byz \wedge Czx \leftrightarrow Czy \wedge Cxz \leftrightarrow Cyz \wedge \forall z'(Dzz'x \leftrightarrow Dzz'y \wedge Dzxz' \leftrightarrow Dzyz' \wedge Dxzz' \leftrightarrow Dyzz'))$$

De acordo com Quine (1986, p. 63), na situação mencionada, “ $x = y$ ” é então definida como a abreviação para uma sentença complexa similar a (Q). A sentença (Q) é a tentativa de expressar a identidade através de

<sup>9</sup> Quine fornece uma demonstração em que o axioma  $(\alpha_2)$  está na forma  $\sim (x = y \wedge (Fx \wedge \sim Fy))$ . Cf. (QUINE, 1986, p. 62-3).

indistinguibilidade de dois objetos  $x$  e  $y$  que satisfazem os predicados da linguagem, mesmo em suas relações com quaisquer outros objetos  $z$  e  $z'$ . A proposta alcança supostamente seu objetivo pelas seguintes razões. De um ponto de vista geral, como (Q) é uma conjunção, ela só poderá ser verdadeira caso cada um de seus componentes sejam verdadeiros. Além disso, trata-se de uma conjunção de bi-implicações. Como bi-implicações são verdadeiras apenas quando seus componentes recebem o mesmo valor de verdade, seja verdadeiro ou falso, caso alguma dessas bi-implicações seja falsa, isso quer dizer que  $x$  ou  $y$  discordam em pelo menos uma propriedade, utilizando o raciocínio contrapositivo da indiscernibilidade dos idênticos, podemos concluir nesse caso que  $x$  e  $y$  são diferentes. Em seu método, Quine utiliza-se implicitamente do princípio recíproco da identidade dos indiscerníveis, nomeadamente, o princípio de *identidade dos indiscerníveis*, segundo o qual se tudo que é verdadeiro acerca de  $x$  também mantém-se verdadeiro sobre  $y$ , então  $x$  e  $y$  são idênticos. Formalizando essa tese em  $LPO^-$ , obtemos ( $L_1$ ):  $\forall x\forall y\forall F(x = y \rightarrow (Fx \leftrightarrow Fy))$ .

Após fornecer seu método de obtenção de um substituto para a identidade, Quine faz a seguinte observação:

Pode acontecer que os objetos pretendidos como valores das variáveis da quantificação não são completamente distinguíveis uns dos outros pelos quatro predicados. Quando isso ocorre, (3) falha em definir identidade genuína. Ainda assim, tal falha permanece inobservável de dentro da linguagem; desse ponto de vista (3) é tão boa quanto a identidade (QUINE, 1986, p. 63)<sup>10</sup>.

Em suma, mesmo que objetos possam satisfazer exatamente os mesmos predicados combinados na forma de (Q), isso não garante que eles sejam de fato idênticos. Por exemplo, se apenas um predicado é usado para definir a identidade, então na esteira de (Q),  $x = y$  deveria ser

<sup>10</sup> "It may happen that the objects intended as values of the variables of quantification are not completely distinguishable from one another by the four predicates. When this happens, (3) fails to define genuine identity. Still, such failure remains unobservable from within the language; (3) is as good as identity from that vantage point."

entendido como  $\forall z(Pzx \leftrightarrow Pzy \wedge Pxz \leftrightarrow Pyz)$ . Se objetos que temos como alvo de nossas quantificações são seres humanos, e  $P$  é o predicado ‘ter o mesmo peso’, então diferentes pessoas com o mesmo peso serão consideradas como a mesma pessoa. Quando  $x$  e  $y$  têm o mesmo peso, qualquer  $z$  que tenha o mesmo peso que  $x$  terá o mesmo peso que  $y$ , e se  $x$  tem o mesmo peso que qualquer  $z$ ,  $y$  também terá. Esse é um caso onde o método proposto por Quine parece falhar em definir a identidade, no entanto, como o próprio Quine observa, do ponto de vista da linguagem, isso é imperceptível. De todo modo, para alguns filósofos muito do que Quine afirma não é tão direto quanto parece.

Quine pretendeu dar conta da identidade por que ela, de certo modo, desafiava sua concepção de ‘verdade lógica’. Ao tentar argumentar em favor da logicidade da identidade, para com isso evitar que ela fosse interpretada de outro modo, Quine forneceu uma série de argumentos, a saber, que identidade é completa, universal e unicamente caracterizável. Entretanto, o autor chegou à conclusão de que não precisaria, de todo, da identidade. Afinal, dado os predicados da linguagem, poderíamos fornecer um substituto que cumprisse o seu papel. Como ressaltamos, o método fornece a indistinguibilidade como substituto da identidade. A seguir são apresentados alguns problemas que parecem afetar a abordagem aqui descrita.

### **Crítica sobre a proposta de Quine**

Quine é acusado de ser imprevidente ao empregar de modo não qualificado o termo ‘identidade’. Basicamente essa é a objeção de Béziau (2013). Em linhas gerais, sua motivação é a de que muito embora os axiomas

$$(\alpha_1): \forall x(x = x)$$

$$(\alpha_2): \forall x \forall y (x = y \rightarrow (Fx \leftrightarrow Fy))$$

sejam considerados axiomas da identidade, não podemos garantir que (=) sempre será interpretado como identidade numérica – a relação que uma coisa tem consigo mesma e nada mais. No que é conhecido como modelos normais para a identidade, (=) é interpretado como a diagonal do domínio D do discurso  $\Delta_D = \{ \langle x, x \rangle : x \in D \}$ . É possível fornecer interpretações que satisfaçam os axiomas ( $\alpha_1$ ) e ( $\alpha_2$ ) mas sem interpretar a identidade como a diagonal<sup>11</sup>. O que se pode garantir, no entanto, é que os axiomas anteriormente listados definem uma relação de congruência, i.e., uma relação de equivalência que é compatível com operações da linguagem de primeira ordem. Com isso em mente o autor defende que

Não é possível concluir que a identidade trivial [identidade numérica] é lógica a partir de (e) [identidade é redutível] ou (d) [identidade é unicamente caracterizável] visto que (e) ou (d) não se aplicam à identidade trivial, tudo o que sabemos é que (e) e (d) valem para identidade de Leibniz. É suficiente chamar de lógica a identidade de Leibniz? (BÉZIAU, 2013, p. 4)<sup>12</sup>.

Segundo Béziau (2013, p. 4-5), o mais adequado seria dizer que, em suas considerações, Quine está falando da identidade de Leibniz (doravante  $=_L$ ) e não da identidade trivial, negando assim que se possa defender a logicidade da identidade a partir das teses defendidas por Quine. Questionamentos interessantes, que podem ser vias a serem explorados surgem aqui. O que aconteceria, no entanto, se a noção de identidade numérica, que ressalta a natureza binária da identidade, fosse um equívoco? Seria suficiente considerar  $=_L$  como constante lógica? Prossequimos com outros contra-argumentos a proposta de Quine.

Wiggins (2001, p. 185) argumenta que o método proposto por Quine faz pouca justiça a evidente univocidade presente na afirmação de que ‘x é o mesmo que y’. A ideia básica é que dadas diferentes teorias, por exemplo,  $T_1$  e  $T_2$ , com diferentes estoques de predicados, a única coisa que podemos garantir é uma noção de indistinguibilidade relativa a cada teoria. Desse

---

<sup>11</sup>Cf. Hodges (2001, p. 64) para um esboço da demonstração.

<sup>12</sup> “It is not possible to conclude that trivial identity is logical from (e) or (d) since (e) or (d) does not hold for trivial identity, all we know is that (e) and (d) hold for Leibniz identity. Is it enough to call Leibniz identity logical?”

modo, haveria uma noção de identidade relativa a  $T_1$  e uma noção de identidade relativa a  $T_2$ . E de acordo com Wiggins, isso levantaria dificuldades ao tentar explicar como falantes competentes com diferentes vocabulários, que, pelo método quineano, deverão fornecer diferentes substitutos para a identidade, compreendem uns aos outros. Além disso, com relação ao método proposto por Quine, Wiggins faz a seguinte observação:

Pois o método de Quine faz qualquer sentido para nós por sua alusão oculta a, e dependência em, uma ideia que é essencialmente de segundo nível. A ideia de segundo nível é a de que a identidade é a relação cuja posse entre  $a$  e  $b$  assegura que toda propriedade de  $a$  é uma propriedade de  $b$  e *vice versa*. O problema com uma abordagem de segundo nível da identidade, contudo, é que ela envolve quantificar sobre todas as propriedades de primeiro nível, incluindo identidade reconstruída e outras incontáveis propriedades que latentemente envolvem a identidade (WIGGINS, 2001, p. 185)<sup>13</sup>.

Em linhas gerais, a questão de Wiggins é ‘*O que cada substituto para identidade tem em comum uns com os outros?*’. Como já mencionamos com a crítica de Béziau, a abordagem fornece relações de congruência. Entretanto, Wiggins argumenta, não podemos expressar em uma linguagem de primeira ordem o que faz com que cada substituto seja de fato uma relação de identidade. Segundo Wiggins, isso só poderia fazer sentido caso apelassem para uma intuição de segunda ordem. Todavia, apelar para intuições de segunda ordem envolveria quantificar sobre propriedades que incluem os substitutos para a identidade e propriedades que alegadamente envolveriam a identidade. Qualificando desse modo um tipo de raciocínio circular. Partimos da tentativa de definir a identidade por meio de predicados da linguagem, mas esse método apenas faria sentido caso pudéssemos explicá-lo recorrendo a intuição de segunda

---

<sup>13</sup> “For Quine’s method makes whatever sense it makes for us by its surreptitious allusion to and dependence upon an idea which is essentially second level. The second level idea is that identity is the relation whose holding between  $a$  and  $b$  ensures that every property of  $a$  is a property of  $b$  and *vice versa*. The trouble with a second level account of identity, however, is that it involves quantifying over all first level properties, including reconstructed identity and countless other properties that latently involve identity.”

ordem, mas isso faz com que quantifiquemos pressupondo a identidade, assim fechamos o círculo. Contudo, essa conclusão pode ser resistida. Primeiramente, Wiggins atribui a Quine o comprometimento com o princípio de ‘indiscernibilidade do idênticos’ formulado em segunda ordem, a saber,

$$(L_1): \forall x \forall y \forall F (x = y \rightarrow (Fx \leftrightarrow Fy))$$

quando, na verdade, Quine reconhece seu comprometimento com uma forma amena da identidade dos indiscerníveis, a saber,  $(L_2) \forall x \forall y ((Fx \leftrightarrow Fy) \rightarrow x = y)$ . Se colocarmos a sentença **(Q)** no escopo dos quantificadores em (L), obtemos o que poderia ser chamado de identidade de Quine (doravante  $=_Q$ ), ou seja,

$$=_Q \forall x \forall y ((Q) \rightarrow x = y)$$

Segundo, a argumentação de Quine tem em vista o caráter lógico da identidade, tentando argumentar o quão próximo ela está da lógica, ao invés de ser tratada como qualquer outro predicado. Não é seu objetivo definir ou eliminar a identidade de uma vez por todas, mas sim fornecer meio tangíveis de lidar formalmente com essa noção dentro uma teoria. O que nos dirige a objeção descrita abaixo.

Se a linguagem que usamos contasse com um número infinito de predicados, seria impossível definir a identidade seguindo os moldes da sentença **(Q)**, e este é um ponto sobre o qual Quine estava ciente. A questão natural a se levantar, contudo, é a de saber quantos predicados são necessários para se obter um substituto adequado para a identidade. Wiggins menciona o seguinte exemplo<sup>14</sup>. Suponha que tivéssemos uma linguagem com três predicados, a saber, ‘x é uma floresta’, ‘x é uma árvore’ e ‘x nasce em y’. Caso adotássemos a metodologia quineana, a sentença

$$(4) \text{ Em toda a floresta nasce mais de uma árvore}$$

<sup>14</sup> O exemplo é atribuído a John Wallace. Cf. Wallace, J. *Philosophical Grammar*. Stanford University Ph.D, 1964.

deveria ser falsa. Isso porque, seguindo a receita fornecida por Quine, o enunciado de identidade ' $x = y$ ' seria expresso, para o presente caso, como

$$(5) \text{Arvore}(x) \leftrightarrow \text{Arvore}(y) \wedge \text{Floresta}(x) \leftrightarrow \text{Floresta}(y) \wedge \\ \forall z(\text{NasceEm}(z,x) \leftrightarrow \text{NasceEm}(z,y) \wedge \text{NasceEm}(x,z) \leftrightarrow \text{NasceEm}(y,z))$$

Note que uma formalização de (4), na lógica de primeira ordem, seguiria os seguintes moldes

$$(4^*) \forall x(\text{Floresta}(x) \rightarrow \exists y \exists z(\text{Arvore}(y) \wedge \text{Arvore}(z) \wedge \text{NasceEm}(y,x) \wedge \\ \text{NasceEm}(z,x) \wedge (\sim y = z)))$$

Nessa formalização, utilizamos ( $\sim y = z$ ) para expressar que o fato de que o número de árvores é maior que um. Desse modo, ao negar que  $y = z$  estamos afirmando que algum dos termos das conjunções em (5) falham. Agora, suponha que  $a$  e  $a'$  sejam árvores que nasçam em uma floresta  $f$ . Usando (5) não é possível diferenciar essas árvores, pois se  $a$  nasce em  $f$ , então  $a'$  também nasce, e vice-versa; se  $a$  é uma árvore, então  $a'$  também o é, e vice-versa etc... Ou seja, (5) continua sendo satisfeita, e, do ponto de vista da linguagem, existe apenas uma árvore em questão. Sem dúvida, uma paisagem quase desértica. Portanto, nessa linguagem cujos únicos predicados são ' $x$  é uma floresta', ' $x$  é uma árvore' e ' $x$  nasce em  $y$ ', a sentença (4) é falsa.

A sugestão apontada por Wiggins para evitar esse problema indesejado é reformular a tese de redutibilidade da identidade de modo a permitir a expansão do estoque de predicados. Entretanto, o seguinte problema surgiria:

[...] tudo é visto depender sobre quais predicados são adicionados para forçar o valor de verdade desejado sob as sentenças envolvendo o predicado construído que é para ser introduzido no lugar de  $=$ . Suponha que as aparências possam apenas ser salvas se são disponibilizados predicados monádicos e poliádicos pressupondo a identidade ou individuação espaço-

tempo ou individuadores de coisas (*thing-individuation*). Se é assim, é sob a presença deles que irá depender o sucesso do método de eliminação (WIGGINS, 2001, p. 185)<sup>15</sup>.

Basicamente, a ideia é a de que se não é possível diferenciar árvores dentro da teoria descrita acima, podemos reformulá-la de modo que árvores sejam diferenciadas. Entretanto, Wiggins defende que se os predicados inclusos nessa teoria reformulada devem incluir predicados que pressupõem a identidade para que aí se possa diferenciar árvores, então será por isso que o método de Quine pode ser considerado satisfatório, pura e simplesmente por depender da identidade. Mas se o método é proposto com vistas a fornecer um substituto para a identidade, não faria sentido dizer que ele alcança seu objetivo de forma satisfatória, quando, na verdade, está pressupondo aquilo que pretende eliminar. Acontece que nessa argumentação ao que parece, Wiggins está usando a noção de *critério de identidade*. Seguindo (MCKEON, 2010, p. 54), dizemos que um predicado *P* fornece um critério de identidade quando é determinado o que conta como um e o mesmo *P*. Se os predicados envolvidos na teoria reformulada incluíssem um predicado que trata da localização espaço-temporal, isso deveria envolver a identidade porque precisaríamos entender árvores como unidades discretas. Entretanto, será que deveríamos incluir no mesmo pacote a noção de identidade enquanto item lógico, da identidade que aparece em critérios de identidade? Outra questão que pode ser tratada em paralelo ao problema aqui descrito.

### Considerações finais

Tomar a identidade como indispensável para a formalização de argumentos e para regimentar noções relevantes para a fundamentação das ciências, pode ser visto como uma via para considerá-la uma noção

---

<sup>15</sup> “[...] everything is seen to depend on what predicates are added in order to force the desired truth-value upon the sentences involving the constructed predicate that is to be introduced in lieu of =. Suppose the appearances can only be saved if monadic and polyadic predicates presupposing identity or place-time individuation or thing-individuation are supplied. If so, it is upon the presence of these that the success of the elimination recipe will depend”.

lógica. Quine ataca essa perspectiva ao formular uma metodologia para que possamos criar um substituto para a identidade. No entanto, como vimos, Béziau (2013) afirma que o método de Quine não trata da identidade, mas sim da identidade de Leibniz, pois, para tratar da identidade, ele deveria dar conta da identidade numérica, a relação binária que uma coisa tem apenas consigo mesma e com nada mais. Wiggins (2001) concorda com a ideia de que o método de Quine apenas capta indistinguibilidade, mas adiciona que o método apenas faria sentido se usasse uma intuição de segunda ordem. Além disso, Wiggins (2001) acrescenta que se um predicado adicionado ao método proposto por Quine envolve algum predicado que forneça critérios de identidade, então o método só é satisfatório por pressupor aquilo que pretende eliminar. Perceba que, nessas abordagens, o operador (=) é sempre entendido como referindo, ou coincidindo, com o que chamamos de identidade numérica. Podemos chamar essa abordagem de *representacionalista* com relação à identidade. O termo é tomado de empréstimo de Peregrin (2012), e é usado, de modo propositalmente vago, para a perspectiva de que nossos termos da linguagem são usados para referir a objetos com os quais nos confrontamos no mundo, ou referem a qualquer realidade extralinguística. Ainda que haja disputas sobre se o método de Quine é ou não satisfatório, ainda assim não estamos em posição de afirmar que a identidade não é eliminável. Na verdade, Quine não foi o único defensor de uma tese nesse espírito. A discussão acima mostra, no entanto, que é preciso cautela ao que podemos inferir sobre a identidade ao tratá-la dentro do nosso arcabouço formal predileto. Apelar para um tratamento não representacionalista da identidade pode ser um caminho que lance luz à discussão.

## Referências

BÉZIAU, J-Y. Quine on identity. *Principia*, v. 7, n. 1-2, p. 1-15, 2013.

BENEVIDES, H. A. **Seria a identidade uma constante lógica? Entre lógica e metafísica.**

2017. Dissertação (Mestrado em Filosofia) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal do Rio Grande do Norte, Natal, 2017.

BUENO, O. Why identity is fundamental. **American Philosophical Quarterly**, v. 51, n. 4,

p. 325-32, 2014.

BONNAY, D. Logicality and invariance. **Bulletin of Symbolic Logic**, 14, 1, p. 29-68, 2006.

DUMMETT, M. **Frege: philosophy of language**. First Edition. S.I.: Harper & Row, 1973.

GENTZEN, G. Investigations into logical deduction. In: SZABO, M. E. (ed.). **The collected papers of Gerhard Gentzen**. S.I.: s.Ed., 1969.

GÓMEZ-TORRENTE, M. The problem of logical constants. **Bulletin of Symbolic Logic**, 8,

1, p. 1-37, 2002.

HINTIKKA, K. J. J. Identity, variables, and impredicative definitions. **Journal of Symbolic**

**Logic**. Association for Symbolic Logic, v. 21, n. 3, p. 225-45, 1956.

HODGES, W. Elementary Predicate Logic. In: Gabbay D. M. & Guentner F. (Eds.).

**Handbook of Philosophical Logic**. Springer Netherlands, p. 1-129, 2001.

MACFARLANE, J. Logical Constants. In: ZALTA, Edward N. (Ed.). **The Stanford**

**encyclopedia of philosophy**, 2015.

MCGINN, C. **Logical Properties: identity, existence, predication, necessity, truth**. Oxford

University Press, 2000.

MCKEON, M. W. **The concept of logical consequence: an introduction to philosophical**

**logic**. Peter Lang Pub., 2010.

NOONAN, H.; CURTIS, B. Identity. In: ZALTA, Edward N. (Ed.). **The Stanford**

**encyclopedia of philosophy**. Stanford, 2014.

PEREGRIN, J. What is Inferentialism? Disponível em

<<http://jarda.peregrin.cz/mybibl/PDFtxt/580.pdf>>. Último acesso: 03.07.2020.

QUINE, W. v. O. On what there is. In: **From a logical point of view**. New York: Harper Torchbooks, 1963, p. 1-19

\_\_\_\_\_. **Philosophy of logic**. [S.l.]: Harvard University Press, 1986.

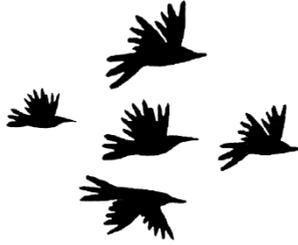
SAVELLOS, E. E. On defining identity. **Notre Dame Journal of Formal Logic**. Duke University Press, v. 31, n. 3, p. 476-84, 1990.

SCHROEDER-HEISTER, P. Popper's theory of deductive inference and the concept of a logical constant. **History and philosophy of logic**, Taylor & Francis, v. 5, n. 1, p. 79-110, 1984.

TARSKI, A. What are logical notions? In: J. Corcoran (ed.). **History and Philosophy of Logic**, 7, p. 143-54, 1986.

WARMBROD, K. Logical constants. **Mind**, v. 108, n. 431, p. 503-38, 1999.

WIGGINS, D. **Sameness and substance renewed**. 2. ed.: Cambridge University Press, 2001.



## Uma nota sobre a relação entre filosofia da mente e filosofia da linguagem

*João de Fernandes Teixeira*<sup>1</sup>

Pode a filosofia da mente ser absorvida pela filosofia da linguagem?

No seu livro *Intentionality*, publicado em 1983, John Searle afirmou que a filosofia da linguagem é parte da filosofia da mente. Embora muitos aspectos da filosofia de Searle não me agradem, neste tópico específico tendo a concordar com ele.

Neste breve capítulo, apresento um contraexemplo pontual à ideia de que a filosofia da mente poderia ser absorvida pela filosofia da linguagem. Examinarei, brevemente, a filosofia da linguagem de Gilbert Ryle, considerado o iniciador da filosofia da mente no século XX.

A história da filosofia da mente começa de forma paradoxal. Ryle (1900-1976) não acreditava que a mente existisse. Ele afirmava que a ideia de que temos uma mente dentro de nós é uma miragem produzida pelo uso inadequado da linguagem cotidiana.

---

<sup>1</sup> João de Fernandes Teixeira é um dos pioneiros da filosofia da mente no Brasil. Bacharel em filosofia pela USP e mestre em filosofia da ciência pela UNICAMP é também PhD pela University of Essex, na Inglaterra. Fez pós-doutorado nos Estados Unidos, sob orientação de Daniel Dennett. Foi colaborador do Instituto de Estudos Avançados da USP e lecionou em várias universidades brasileiras, como a UNESP, a UFSCar e a PUC-SP. Publicou 18 livros na área de filosofia da mente e ciência cognitiva. Mantém a página Filosofia da Mente no Brasil no Facebook (<https://www.facebook.com/filosofiadamentenobrasil/>).

Como bom discípulo de Ludwig Wittgenstein (1889-1951), Ryle sustentava que muitos de nossos problemas filosóficos eram ilusões criadas pela nossa própria linguagem, que sempre nos enredaria em seus inúmeros labirintos. Nesse sentido, o problema mente-cérebro era, para ele, um falso problema gerado por uma dessas ilusões.

Em seu livro *The Concept of Mind*, publicado em 1949, Ryle busca a desconstrução da ideia de mente. Ele apresenta duas estratégias para realizar essa tarefa, ambas derivadas da investigação sobre a linguagem. A primeira é a demarcação entre a linguagem física e a mental, que produz as transgressões categoriais. Uma vez identificadas as transgressões categoriais, o problema mente-cérebro não seria resolvido, nem a favor dos materialistas, nem dos dualistas. Ele simplesmente seria dissolvido, pois teria ficado evidente que ele não passa de um pseudoproblema. A segunda estratégia, derivada da ideia de transgressão categorial é a ideia de que a nossa linguagem produz uma concepção equivocada de mente, o *fantasma na máquina*, que poderia ser exorcizado por uma terapia linguística. Essa terapia consiste em identificar alguns nódulos de nossa linguagem onde ocorrem as transgressões categoriais.

Quando digo “minha mente está cansada hoje” estou tomando “mente” como um objeto e não como um conceito. Estou aplicando o vocabulário físico ao mental, pois meus músculos podem ficar cansados, mas minha mente, não. O mesmo vale para a sentença “Esses dados confundem a cabeça do telespectador”. Nesse caso, a palavra “mente” está sendo indevidamente substituída por algo material, “cabeça”. Isso poderia induzir a falsa ideia de que a mente é algo material, ou seja, estaríamos aplicando o vocabulário físico ao mental. Essa passagem despercebida entre palavras que se aplicam ao físico e ao mental teria, progressivamente, levado ao aparecimento da ideia da mente como uma substância, o fantasma na máquina.

O fantasma na máquina também surge dos equívocos da linguagem cotidiana, no caso, em um processo de falsa expansão ontológica que não passa de uma expansão da linguagem. Ele seria mais um caso de

transgressão categorial, no qual a linguagem induz a formação da ideia de que o mental teria características do físico, ou seja, a substancialidade.

O fantasma, para Ryle, seria a alma e a máquina, o corpo. A doutrina oficial relaciona essas duas partes por meio de uma relação causal inexplicada, que herda do cartesianismo o problema de saber como uma substância imaterial causaria os movimentos do corpo. Como resposta a esse problema, Ryle argumenta que a mente humana não é uma substância, mas uma palavra, um conceito.

Em um exemplo clássico, Ryle ilustra como uma ideia equivocada de mente, que resulta de transgressões categoriais acaba se formando na linguagem cotidiana. Ele imagina que um dia alguém te convida a conhecer a Universidade de São Paulo. Você convida essa pessoa a entrar no seu carro e começa a dar uma longa volta, parando na frente de cada edifício e dizendo: “este é o prédio da filosofia”, “este é o prédio da psicologia” e assim por diante. No fim do dia, você teria mostrado a essa pessoa todos os prédios do campus. Imagine agora se essa pessoa, ao descer do carro e se despedir, agradecendo a gentileza, dissesse: “Muito obrigado, você me mostrou tantos prédios! Mas ainda não me mostrou a Universidade de São Paulo”.

Certamente essa última frase causaria perplexidade. Por acaso, você já não teria mostrado a Universidade de São Paulo para ela? Afinal, o que será que ela está querendo *ver*? Será que ela pensa que a Universidade de São Paulo deve ser um outro prédio para além dos prédios de todas as escolas que você já mostrou? Ou será que ela pensa que, para além de todos esses outros prédios, deve existir *alguma outra coisa* que seja a Universidade de São Paulo?

A situação ilustra o que constitui, para Ryle, um dos principais equívocos da filosofia da mente. Inspirados pela tradição cartesiana, os filósofos da mente começaram a supor que, para além de todo o conjunto de comportamentos e disposições que observamos nas pessoas, existiria algo a mais, algo como uma substância subjacente a todas essas manifestações, da mesma maneira que a pessoa que você levou para

conhecer a Universidade de São Paulo, mas não conseguiu vê-la. Da mesma forma que a Universidade de São Paulo se esgota no conjunto de prédios que a compõem, a mente se esgota no conjunto de comportamentos e disposições manifestados pelas pessoas.

Supor que existe algo a mais do que isso é o equívoco que Ryle aponta. Um equívoco que leva a supor que, para além das partes, deve haver algo inexplicito a elas, algo que os cartesianos identificariam como uma substância imaterial. Supor a existência dessa substância é incorrer em uma ilusão; a ilusão originária que leva a postular a existência de um fantasma na máquina. Um fantasma que, uma vez postulado, leva a todas as dificuldades e problemas insolúveis acerca da natureza da mente e de sua relação com o cérebro e com o mundo físico.

O fantasma na máquina nos faz confundir um conceito com uma coisa. Ou confundir um conceito com algum tipo de substância que deveria existir em algum lugar. Ryle nos diz que a mente não é nada além de um conceito que utilizamos para designar um conjunto de comportamentos e disposições. Alguns tipos de conceitos, como seria o caso do conceito de mente, designam relações e não entidades. Relações que não levam a nada palpável, como, por exemplo, a ideia de “quarta-feira”, que não designa nada tangível, mas apenas uma relação que estipulamos entre os diversos dias da semana.

Curiosamente, um contemporâneo de Ryle, o psicólogo americano B.F. Skinner defendeu posições muito semelhantes. Mas Skinner é citado apenas uma vez no *The Concept of Mind*.

\*\*\*

Será o projeto de Ryle filosoficamente viável? Será que a crítica da linguagem é suficiente para dissolver os principais problemas da filosofia da mente?

Penso que uma das questões mais importantes que surge da leitura do texto de Ryle é saber se a noção de transgressão categorial dissolve o

problema mente-cérebro. Até que ponto a análise linguística pode dar conta do problema central da filosofia da mente? Está fora do meu alcance fornecer uma resposta a essa questão. No entanto, outros filósofos da mente já apontaram que a distinção entre o vocabulário físico e o mental nem sempre é nítida e que, em alguns casos, ela é nebulosa. São contraexemplos ao projeto de Ryle e não sabemos, até agora, o quanto eles são comprometedores.

Não faltaram críticas, como as de Dennett (1969), para a proposta de demarcação da linguagem defendida por Ryle. Dennett argumenta que essa demarcação não pode ser precisa e que em muitas instâncias não há separação nítida entre o físico e o mental, mas, ao contrário, uma porosidade.

Tomemos, por exemplo, as noções de distância e de *medida* de distâncias, tais como quilômetros e metros. Será possível situar a idéia de quilômetro no vocabulário físico? Ou devemos situá-la no vocabulário mental? Podem os quilômetros que existem entre a Terra e a Lua serem identificados com algo *físico* do mundo? Certamente que não. Mas, por outro lado, não é possível conceber a distância a não ser como algo físico. Quilômetros e outras medidas, como, por exemplo, graus centígrados, teriam uma existência tênue entre o físico e o mental – uma existência mais tênue do que aquela dos pensamentos, crenças e desejos. O mesmo ocorre com expressões nas quais aparecem os advérbios “perto” ou “distante” ou “próximo” e “longe”. Esses advérbios designam uma situação física, mas só fazem sentido se forem interpretados subjetivamente, isto é, se alguém achar que está “perto” de algo ou “próximo” de algo. Novamente, temos uma situação-limite na qual o sentido de uma expressão é adquirido precisamente pelo fato dela transitar entre o vocabulário físico e o vocabulário mental.

Certamente não podemos borrar tinta num pensamento, numa crença ou em um desejo, mas podemos dizer que um pensamento ocorre na minha cabeça – e ocorrer no espaço ou em algum lugar certamente

constitui uma propriedade física. Uma dor pode ser intensa, da mesma forma que afirmamos que uma chama de fogão é intensa.

Entretanto, a análise da linguagem dificilmente poderia classificar como transgressões categoriais as afirmações de que o pensamento ocorre na minha cabeça, que a dor é intensa ou que um desejo embrulha meu estômago: estas sentenças não só fazem sentido como o fazem precisamente por transitar entre o físico e o mental. O resultado é inverso no caso dos quilômetros e outras distâncias ou medidas, nos quais a análise da linguagem reverte nossa tendência habitual de situá-los do lado do vocabulário físico.

Um outro exemplo interessante é a análise do termo “voz”. Devemos situar vozes no vocabulário físico ou no vocabulário mental? Quando afirmo “ouço uma voz” ou “perdi minha voz” ou mesmo “ele ouve vozes” será que posso situar “voz” em domínios distintos? Que sentido tem tratar “voz” como coisa física quando afirmo “perdi minha voz”? Terá sentido a afirmação “perder a voz” entendida como perder um *objeto físico*? Por outro lado, a idéia de voz como coisa física não pode ser abandonada quando gravo minha voz numa fita magnética e a mando para um amigo em Londres. Será então “a voz que perco” uma metáfora que significa, na realidade, a perda temporária de uma disposição? Neste caso, para resolver esta dificuldade teríamos apenas de fazer um levantamento dos diferentes sentidos da palavra “voz”: haveria uma alternância entre o sentido físico e o sentido mental. Mas isto rapidamente leva a paradoxos, pois, se assim fosse, a voz que eu emito não poderia ser a voz que eu perdi ontem; uma seria física e outra seria mental. Talvez o mesmo se aplicasse às vozes que ouço quando tenho um surto psicótico: não *penso* vozes, *ouço* vozes, e como poderia saber se elas são físicas ou mentais?

Mas há uma outra objeção. Será que podemos prescindir do fantasma na máquina? Ou, em outras palavras, será possível prescindir do conceito de mente?

Quando me olho no espelho, não vejo minhas partes como ocorreu no caso do turista que viu a Universidade de São Paulo. Vejo o JT (João Teixeira). Será que o JT é apenas um conceito?

Um conceito é insuficiente para conceber o que é uma pessoa. São as pessoas que pensam e não apenas partes delas. Da mesma forma, são as pessoas que cometem crimes e não partes delas reunidas e concebidas como um conceito unificador. Se raciocinarmos de maneira análoga a relação entre prédios e a Universidade de São Paulo, como propõe Ryle, eu poderia afirmar que não fui eu quem cometeu determinado crime, mas meu cérebro, meus braços e mãos. Mas isso seria um contrassenso. Conceitos não cometem crimes. Pessoas cometem crimes. Por isso, uma pessoa não pode ser apenas um conceito.

Será que a ideia de um conjunto de comportamentos e disposições é suficiente para explicar a ação? Será possível explicar como comportamentos e disposições ocorrem formando uma sequência sem introduzir alguma ideia de auto-organização entre organismo e ambiente?

Não sabemos o que é o eu nem tampouco se ele é uma substância. Mas temos uma identidade pessoal cuja existência não podemos negar. Ela pode ser considerada uma ficção, pode ser remodelada ou alterada, mas sempre posso referir a mim mesmo como sendo uma pessoa. Temos um RG, um CPF, que pode ser desconhecido por todos os que convivem conosco, mas que levará a alguém, a uma pessoa, viva ou morta, que sempre será (ou foi) algo mais do que apenas um conceito. É nesse ponto que discordo da suposição de Ryle de que JT poderia ser apenas um conceito equivocado, construído a partir da linguagem.

\*\*\*

Ryle foi um dos filósofos mais importantes do movimento que ficou conhecido como a “virada linguística” da filosofia do século XX. Para esses filósofos, a reflexão sobre a linguagem deveria preceder a discussão de problemas metafísicos, incluindo as questões da filosofia da mente. A

importância da análise da linguagem nunca foi negada pelos filósofos da mente. No entanto, elegê-la como tema primordial da filosofia da mente é uma posição que deve ser questionada, pois, como vimos, os objetos da filosofia da mente extrapolam sua enunciação linguística.

Ryle parece ter esquecido que a linguagem é uma manifestação da mente. Se a linguagem é um produto da mente como ela poderia ser utilizada para dissolver a ideia de mente? Uma análise linguística do problema mente-cérebro corre o risco de se tornar circular. É nesse sentido que não concordo com a proposta de que a filosofia da mente pode ser absorvida pela filosofia da linguagem.

No final da década de 1940, quando o livro de Ryle foi publicado, ainda não existiam computadores de mesa. Mas, depois que eles se popularizaram, 40 anos depois, passamos a empregar o vocabulário psicológico para descrever algumas de suas operações. Falamos que o computador “não reconhece” um dispositivo, que ele “não consegue ler um arquivo”. Se a hipótese linguística da formação da mente defendida por Ryle estiver correta, deveríamos atribuir aos computadores um fantasma na máquina. Ironicamente, não teríamos mais apenas um fantasma na máquina como supôs Ryle. Eles variariam de acordo com o fabricante e a marca dos computadores. Haveria milhões de fantasmas, pois poderíamos atribuir um para cada máquina cujo comportamento pudesse ser descrito por meio do vocabulário psicológico humano. Será que estaríamos cometendo uma transgressão categorial todas as vezes que usássemos esse vocabulário para descrever as operações do computador?

\*\*\*

Este capítulo está longe de responder de forma definitiva à questão sobre a possibilidade de a filosofia da mente ser assimilada pela filosofia da linguagem. Analisei apenas um fragmento desse debate ao abordar a obra de Gilbert Ryle, mais especificamente *The Concept of Mind*. Penso que para responder afirmativamente essa questão seria necessário demonstrar

que o mental pode ser inteiramente compreendido como um construto linguístico. Creio ter argumentado, pelo menos parcialmente, que isso não é possível.

## **Referências**

DENNETT, D. **Content and Consciousness**. London: Routledge & Kegan Paul, 1969.

RYLE, G. **The Concept of Mind**. New York: Barnes & Noble, 1949.

SEARLE, J. **Intentionality: An Essay in the Philosophy of Mind**. Cambridge: Cambridge University Press, 1983.



## Existem sensações inconscientes? Um exame da teoria das sensações de David Rosenthal

*Tárik de Athayde Prata*<sup>1</sup>

### 1. Introdução

De acordo com uma perspectiva altamente influente na história da filosofia e dotada de um forte apelo intuitivo (como veremos a seguir), a consciência é uma propriedade *essencial* dos fenômenos mentais. Esse modo de ver é amplamente identificado com o pensamento de René Descartes (cf. ROSENTHAL, 1986, p. 331; ARMSTRONG, 1997, p. 721; ARMSTRONG, 1999, pp. 14-16; SEARLE, 1989, p. 203; SEARLE, 1992, p. 164; SEARLE, 1997, p. 236; SEARLE, 2004, p. 13; SEARLE, 2015, p. 202; MARQUES, 2017, p. 67-68) e, de fato, o autor do *Discurso do método* caracterizou o pensamento – para ele o atributo essencial da mente – como “tudo quanto está de tal modo em nós que somos disso imediatamente conscientes.” (DESCARTES, 1979, p. 169 [AT, VII, p. 160; AT, IX, p. 124;

---

<sup>1</sup> Doutor em Filosofia (2007) pela Ruprecht-Karl Universität Heidelberg (Alemanha), orientado pelo Professor Andreas Kemmerling. Foi bolsista de Desenvolvimento Científico Regional (CNPq/Funcap) no Departamento de Filosofia da UFC (2007 a 2009), Professor Adjunto do Departamento de Filosofia da Universidade Federal de Sergipe (UFS) (2009 a 2011) e editor da revista Cadernos UFS de Filosofia. Atualmente, é Professor Associado, nível 2, no Departamento de Filosofia da Universidade Federal de Pernambuco (UFPE), onde foi editor da revista Perspectiva Filosófica (2013), vice-coordenador do Mestrado em Filosofia (2014) e coordenador do Mestrado (2015 a 2018).

CSM, II, p. 113]<sup>2</sup>), o que exclui a possibilidade de fenômenos mentais desprovidos de consciência<sup>3</sup>. Contudo, é bem conhecido que esse modo de ver é bastante anterior a Descartes, pois milênios se passaram antes que a tese de que existem fenômenos mentais inconscientes fosse defendida na cena filosófica (cf. BRENTANO, 1924, p. 144; BRENTANO, 1995, p. 103; BARGH & MORSELLA, 2008, p. 73).

Entretanto, desde a virada do século XX, a aceitação do inconsciente se tornou majoritária, seja por força da obra de Sigmund Freud (cf. SEARLE, 1992, p. 151; SEARLE, 1997, p. 218; GENNARO, 1996, p. 6) seja, mais recentemente, devido às contribuições das ciências cognitivas (cf. TALVITIE, 2009, p. 9; BOAG, 2015, p. 240; SEARLE, 2015, p. 203). E, na verdade, a aceitação do inconsciente se mostra bastante intuitiva, se considerarmos que, a cada instante, possuímos conhecimentos, memórias e preferências em número muito superior ao que pode ocorrer em nossa consciência (cf. ROSENTHAL, 1986, p. 334).<sup>4</sup>

Mas a intuição cartesiana – da consciência como essencial para os fenômenos mentais (cf. ROSENTHAL, 1986, p. 331) – preserva sua força sobre nossa compreensão da mente, em especial, no que diz respeito a um dos mais importantes tipos de fenômenos mentais: as *sensações* (sejam corporais ou perceptivas, bem como outros fenômenos marcados por caráter sensorial – cf. a nota de rodapé nº 6, a seguir). A interação causal de nosso aparato perceptivo com nosso ambiente circundante envolve a apreensão de estímulos que gera *qualidades* vivenciadas conscientemente,

---

<sup>2</sup> Entre colchetes se encontra a referência segundo o sistema que é o padrão internacional nos estudos cartesianos. Os números romanos dão as referências nos volumes em latim (volume VII) e em francês (volume IX) da edição completa das obras de Descartes feita por Charles Adam e Paul Tannery (AT), bem como da edição em inglês (volume II) preparada por John Cottingham, Robert Stoothoff e Dugald Murdoch (CSM).

<sup>3</sup> “Nenhum pensamento pode existir em nós do qual nos não estejamos conscientes no exato momento em que ele existe em nós” (cf. “Quartas respostas”, In: AT, VII, p. 246; AT, IX, p. 190; CSM, II, p. 171). Essa passagem citada não se encontra disponível na edição brasileira das meditações que foi consultada para a redação do presente texto.

<sup>4</sup> De acordo com Franz Brentano: “Naturalmente eles [os filósofos] estavam bem familiarizados com o fato de que se pode possuir um tesouro de conhecimentos adquiridos, sem pensar neles; mas de modo totalmente correto eles concebiam os mesmos como disposições para certos atos de pensar, como também concebiam o caráter adquirido como disposição para certos afetos e atos de vontade, mas não como conhecimento e consciência” (BRENTANO, 1924, p. 144; BRENTANO, 1995, p. 103). Aqui Brentano adere a uma concepção disposicional do inconsciente, que é bastante problemática (cf. PRATA, 2018), mas que não será discutida no presente trabalho.

qualidades simples (de natureza visual, auditiva, olfativa etc.), que se articulam na construção de *percepções* complexas de objetos e estados de coisas no ambiente<sup>5</sup>.

O caso é que tais qualidades constitutivas das sensações parecem estar necessariamente ligadas ao nosso efetivo *sentir*, de maneira que uma sensação não sentida nos aparece como uma impossibilidade. Como colocou o filósofo escocês Thomas Reid (1710-1796): “Quando [uma sensação] não é sentida, ela não é. Não há diferença entre uma sensação [*sensation*] e o sentir [*feeling*] dela, eles são uma e a mesma coisa” (Reid, *Essays*, apud Rosenthal, 1986, p. 358, nota n° 19). No século XX, essa maneira de entender as sensações foi defendida, por exemplo, por alguém da estatura de Saul Kripke (1940- ), que afirmou: “para uma sensação ser *sentida* [*felt*] como dor é para ela *ser* uma dor” (Kripke, *Identity and Necessity*, apud ROSENTHAL, 1986, p. 347), de maneira que, inversamente: “para [algo] existir sem ser *sentido como dor* é para ele existir sem que *exista nenhuma dor*.” (Kripke, *Naming and Necessity*, apud. ROSENTHAL, Op. cit., p. 347)<sup>6</sup>.

A identificação da dor com a sua “qualidade fenomenológica imediata” (KRIPKE, 2001 [1972], p. 152) desempenha um importante papel na crítica de Kripke à teoria da identidade mente-cérebro na filosofia da mente, pois a qualidade fenomenológica dessa sensação nos aparece como algo *qualitativo* e, assim, completamente distinto de qualquer propriedade física (que é objetiva, desprovida de qualidades), de tal modo que permanece *concebível* para nós que exista dor sem que ela esteja

---

<sup>5</sup> “Tanto em nível biológico quanto psicológico de análise, faz-se uma distinção entre sensação e percepção. Em nível psicológico, as *sensações* são experiências associadas com estímulos simples (como uma luz vermelha piscando), e a *percepção* envolve a integração e interpretação significativa das sensações (por exemplo: “É um carro de bombeiros”)” (ATKINSON, R. et al., 2009, p. 110; ATKINSON, R. et al., 2002, p. 133).

<sup>6</sup> Mais recentemente, o influente filósofo Michael Tye expressou o mesmo tipo de convicção: “Entre nossos estados mentais conscientes, alguns são inerentemente conscientes. Isto significa que alguns de nossos estados mentais não podem *falhar* em ser conscientes. Para cada um de tais estados mentais, há uma perspectiva *subjetiva* que o acompanha. Essa perspectiva é conferida ao sujeito simplesmente por ele ou ela estar experimentando o estado mental. Ela é capturada na linguagem cotidiana pela expressão “como é” [*what it's like*]. Há algo como é sentir subjetivamente uma coceira, cheirar ovos podres, provar um limão, se sentir exaltado” (TYE, 2007, p. 23). Tye classifica os estados mentais “fenomenalmente conscientes”, que são (segundo ele) inerentemente conscientes, nas seguintes categorias: (1) experiências perceptivas; (2) sensações corporais; (3) Emoções e (4) Estados de humor.

correlacionada a um processo físico (como a célebre “estimulação das fibras C”), bem como permanece concebível que exista um processo físico que não está correlacionado a nenhuma dor. Sendo assim, a asserção de identidade entre a dor e um processo cerebral se mostra como uma asserção *contingente*, de modo que (dado que ela é construída com dois *designadores rígidos* – designadores que designam seus referentes em todos os mundos possíveis) ela tem que ser uma asserção *falsa*, já que quaisquer asserções de identidade verdadeiras, se forem construídas com designadores rígidos, tem que ser verdadeiras em *todos* os mundos possíveis<sup>7</sup>.

É interessante notar que, ao articular essa argumentação, Kripke está retomando, declaradamente, o principal argumento de Descartes a favor da distinção real entre a alma e o corpo do homem<sup>8</sup>, resultando no célebre dualismo cartesiano de substâncias, pois, embora Kripke pareça se inclinar na direção de um dualismo de propriedades, fica claro que a consideração da consciência como essencial para os fenômenos mentais está correlacionada a uma posição ontológica dualista<sup>9</sup>.

Contudo, somos, de fato, obrigados a aceitar que sensações são necessariamente conscientes? É claro que recusar essa necessidade parece bastante contraintuitivo, afinal, como poderia haver uma sensação que não é sentida? Mas a verdade é que existem motivações teóricas para recusar essa visão cartesiana, já que sua aceitação gera problemas de grande magnitude para a filosofia da mente. Se todos os fenômenos mentais (sejam sensações e percepções, sejam atitudes proposicionais como crenças e desejos ou mesmo estados emocionais) são essencialmente conscientes, então essa propriedade de ser consciente se torna misteriosa,

---

<sup>7</sup> Sobre o argumento de Kripke, cf. Maslin (2001, p. 96-100), Maslin, (2009, p. 97-100), bem como Prata (2009) e Prata (2011).

<sup>8</sup> Cf. Descartes (1979, p. 79-80 [AT, VII, p. 13; AT, IX, p. 9-10; CSM, II, p. 9]), bem como Descartes (1979, p. 134 [AT, VII, p. 78; AT, IX, p. 62; CSM, II, p. 54]).

<sup>9</sup> Descartes, diferente de Kripke, não baseava seu argumento para o dualismo em propriedades qualitativas, pois as sensações eram vistas por ele como uma forma inferior de fenômenos mentais, mas ele considerava a consciência como uma propriedade intrínseca de atitudes proposicionais, assim como Kripke considera o aspecto qualitativo uma propriedade intrínseca das sensações, e o aspecto qualitativo parece, intuitivamente, ser necessariamente consciente.

pois não temos como explicá-la em termos psicológicos, pois uma tal explicação seria baseada em algum fenômeno mental, mas esse fenômeno, já seria, em princípio, um fenômeno consciente, de modo que a explicação da consciência seria feita em termos de um fenômeno consciente, tornando-se *circular* (cf. ROSENTHAL, 1986, p. 330; ROSENTHAL, 1997, p. 735; ROSENTHAL, 2017, p. 154-55).

Por essa razão, o filósofo americano David Rosenthal propõe uma teoria na qual a consciência (enquanto uma *propriedade* de estados mentais) é uma propriedade contingente, devendo a sua instanciação em um dado fenômeno psíquico a algum tipo de *relação* entre esse fenômeno e algum *outro* fenômeno mental. Portanto, Rosenthal propõe um tipo peculiar de “teoria de ordem superior”<sup>10</sup>, na qual a consciência como uma propriedade de um estado mental de primeira ordem é explicada em termos da ação de um outro estado mental que é *acerca* (ou *a respeito*) do primeiro, e esse estado mental de ordem superior é concebido como um *pensamento*<sup>11</sup>. Em tal teoria, a consciência é uma propriedade *contingente* de estados mentais, pois o fato de um estado mental se encontrar na relação adequada com algum outro estado mental não é constitutivo desse estado, pois todo estado mental pode continuar a existir quando não se encontra em tal relação.

O caso é que uma teoria como essa implica que *todos* os fenômenos mentais podem existir de forma inconsciente, o que obriga Rosenthal a argumentar a favor da tese (contraintuitiva) de que *sensações* podem existir inconscientemente. Na verdade, Rosenthal precisa não só (a) viabilizar a aceitação da existência de sensações e aspectos qualitativos

---

<sup>10</sup> A caracterização de “superior” não tem um caráter valorativo, mas diz respeito apenas ao fato de que temos estados mentais que são *acerca* de outros estados mentais. Como esclarece Rosenthal: “Esses pensamentos não de ordem superior apenas no aspecto de terem conteúdo intencional que é *sobre* [about] outros estados psicológicos” (ROSENTHAL, 2008, p. 835, grifo meu).

<sup>11</sup> Um aspecto importante da teoria da consciência de Rosenthal são as três distinções conceituais que ele propõe a respeito da consciência, a saber: a distinção entre (1) consciência de *criatura* e consciência de *estado*, a distinção entre (2) consciência *não introspectiva* e consciência *introspectiva*, e a distinção entre (3) consciência *intransitiva* e consciência *transitiva*. A teoria dos “pensamentos de ordem superior” é concebida por ele como uma teoria a respeito da consciência intransitiva de estado (tanto a introspectiva quanto a não introspectiva). Para uma discussão dessas três distinções, cf. Prata (2017a) e Prata (2017b).

inconscientes, mas precisa também (b) mostrar que sua teoria é capaz de explicar satisfatoriamente as características das sensações conscientes. E ele, de fato, se esforça para mostrar isso. Mas, no presente trabalho, por entender que ele constitui uma questão mais fundamental, irei me concentrar no ponto (a), de deixando o ponto (b) para estudos futuros. Antes de se verificar se a filosofia da mente de Rosenthal é capaz de dar conta das características fenomenais das sensações conscientes (b), é preciso verificar se ele é bem sucedido em fundamentar a crença em sensações inconscientes (cuja existência é um compromisso básico de sua teoria).

O presente trabalho não irá se ocupar, em grandes detalhes, das consequências da aceitação de sensações e qualidades inconscientes para o argumento kripkeano a favor de um dualismo de propriedades (embora algumas dessas consequências sobre outros argumentos anti-fisicalistas sejam rapidamente discutidas a seguir – cf. a seção 3). O foco principal estará direcionado para o item (a) mencionado acima, ou seja, para os *argumentos* de Rosenthal a favor dessa aceitação, e o objetivo do texto é avaliar em que medida esses argumentos são convincentes.

A importância disso reside no fato de que a capacidade de fundamentar a aceitação de sensações inconscientes é decisiva para que a teoria de Rosenthal seja capaz de dar conta do aspecto *fenomenal* da consciência, que é um dos aspectos mais importantes no debate contemporâneo. Ou seja, caso a aceitação da existência de sensações inconscientes se mostre inviável, então a teoria dos pensamentos de ordem superior parecerá estar diante de um *contraexemplo* que contradiz uma de suas assunções fundamentais, a saber, a assunção de que *todos* os fenômenos mentais (tanto intencionais quanto sensoriais) podem existir de forma inconsciente.

No restante do presente texto, após uma discussão de alguns dos argumentos de Rosenthal para a existência de sensações inconscientes (seção 2), serão discutidas as linhas fundamentais para a teoria das sensações que ele desenvolve para explicar essa existência inconsciente, a

saber, a “teoria do espaço de qualidades” (seção 3), e, por fim, será feita uma breve avaliação dos argumentos a favor da existência de sensações inconscientes (considerações finais).

## 2. Alguns argumentos de Rosenthal

De fato, parece bem mais fácil aceitar a existência de *estados intencionais* inconscientes (como crenças, desejos e intenções) do que aceitar que podemos possuir *estados sensoriais* dos quais não temos consciência. Por um lado, parece não apenas bastante razoável, mas, na verdade, inquestionável, que possuímos diversas crenças, desejos e intenções que permanecem inconscientes durante a maior parte do tempo. Considerando que tenho a minha frente, enquanto escrevo estas palavras, uma semana inteira de trabalho, ao longo da qual terei de cumprir uma série de compromissos, parece inegável que eu tenho, neste momento, diversas intenções (necessárias para praticar as ações que serão exigidas em meus compromissos) nas quais não estou pensando neste momento (de modo que elas se mostram como intenções inconscientes). Do mesmo modo, considerando, por exemplo, que passei aproximadamente dez dias de férias em Minas Gerais, no passado recente, parece inegável que eu possuo uma série de crenças sobre as diversas cidades que visitei – Belo Horizonte, Ouro Preto, Brumadinho, Tiradentes, São João del-Rei etc. –, crenças que são de um número muito maior do que eu posso pensar conscientemente a cada instante – de modo que elas se mostram como crenças inconscientes – (cf. ROSENTHAL, 1986, p. 334).

Por outro lado, que sentido parece haver em dizer que tenho, agora, uma dor que não sinto? Ou que posso ter a *sensação* de um calor ambiente apesar de eu não *sentir-lo*? Tais hipóteses apenas reforçam a intuição (já elaborada por Reid, Kripke e Tye) de que não podem existir sensações desprovidas da propriedade de serem conscientes.

Porém, em sua argumentação a favor dessa possibilidade, Rosenthal procura mostrar que essa intuição pode ser concebida como uma mera

consequência de nossos *interesses* pragmáticos. De acordo com ele, podemos explicar nossa tendência a associar consciência e qualidades sensoriais se observarmos que temos grande interesse em nossas sensações corporais conscientes e nas de outros, porque elas tem uma grande utilidade como indicadores de bem estar (cf. ROSENTHAL, 1986, p. 348; ROSENTHAL, 1991, p. 18; ROSENTHAL, 1997, p. 732; ROSENTHAL, 2017, p. 149), além do fato de que, evidentemente, não temos nenhum conhecimento de sensações das quais não tenhamos consciência (ROSENTHAL, 1991, p. 17; ROSENTHAL, 1997, p. 732; ROSENTHAL, 2017, p. 148). Por outro lado, a situação dos estados intencionais é diferente, pois frequentemente é bastante útil saber quais são os pensamentos, emoções e percepções de alguém, mesmo quando essa pessoa não está ciente desses estados (cf. ROSENTHAL, 1986, p. 348), pois isso pode ser de grande utilidade para a nossa compreensão de suas ações. Nas palavras de Rosenthal:

Quanto menos útil é saber sobre um tipo particular de estado mental mesmo quando a pessoa não está ciente dele, mais forte é a nossa intuição de que esse tipo de estado mental tem que ser consciente. Essa correlação é uma evidência reveladora de que, mesmo no caso de dores e outras sensações somáticas, a ideia de que ser mental implica ser consciente é apenas um reflexo de nossos interesses usuais, e não um assunto do significado de nossas palavras ou da natureza do mental ele mesmo (ROSENTHAL, 1986, p. 348-9).

Essa alusão ao significado das palavras que usamos para designar os nossos fenômenos mentais já traz à tona um outro assunto que parece reforçar a intuição de que sensações tem de ser necessariamente conscientes, já que certas expressões (em seu próprio significado) implicam consciência. Por exemplo, dizer que algo “faz sofrer” (*hurt*) já implica uma ciência (*awareness*) do sofrimento, e também parece que não se pode *dizer* que alguém está com dor a menos que essa pessoa *saiba* que tem dor (cf. ROSENTHAL, 1986, p. 349). Certas expressões, que inclusive são bastante utilizadas no debate filosófico – por exemplo: “como uma

sensação é” (*what a sensation is like*) ou “como se sente uma sensação” (*how a sensation feels*) – reforçam a impressão de que estados sensoriais implicam consciência, já que essas expressões se referem tanto à qualidade sensorial quanto à nossa ciência dessa qualidade, parecendo misturá-las (ROSENTHAL, 1986, p. 349).

Mas Rosenthal entende que o nosso uso de expressões como essas não prova nada a respeito da suposta necessidade de sensações possuírem consciência, pois essas expressões são usadas em situações nas quais estamos cientes das sensações, de modo que seu uso não exclui a possibilidade de que essas sensações (com suas qualidades específicas) continuem a existir durante o tempo em que não estamos cientes delas. Ele considera perfeitamente plausível que as diferentes sensações (sejam visuais, auditivas, olfativas etc.) preservem suas qualidades distintivas e continuem a se assemelhar e a diferir entre si enquanto permanecem inconscientes da mesma maneira que se assemelham e diferem quando estão conscientes.

Isso seria comprovado por diversas situações cotidianas, nas quais, por exemplo, uma dor de cabeça ou um pequeno desconforto corporal se integra ao nosso fluxo de consciência de maneira *intermitente*, devido a distrações que nos levam a prestar atenção em outras coisas. Em tais situações, nós não tendemos a pensar que a dor ou o desconforto deixou de existir durante o tempo em que estávamos pensando em outra coisa, mas sim que a dor ou o desconforto existem continuamente, independentemente de estarmos cientes deles ou não, de maneira que alegar que estados inconscientes não podem ter qualidades sensoriais parece a Rosenthal ser uma adesão tácita à definição cartesiana da mente. De acordo com ele:

Termos tais como ‘sentimento’ trazem de fato esta implicação de consciência; uma dor sentida é obrigatoriamente uma dor consciente. Isto também é verdade se algo faz sofrer [*hurting*], e talvez até mesmo se alguém estiver simplesmente com dor. Mas estas situações não são o mesmo que ter uma dor, ou o mesmo que a existência de uma dor. Se estamos intermitentemente

desatentos a uma dor por estarmos distraídos dela, sentimos a dor apenas de maneira intermitente; o mesmo ocorre com seu caráter sofrido [*hurting*] e com nossa situação de estar com dor. Ainda assim, é natural falar que se teve uma dor que persistiu durante o dia todo, e mesmo falar que não se esteve o tempo todo ciente dessa dor. Isto nos fornece evidência de que o senso comum admite a existência de dores inconscientes. Sentir dores e possuí-las parecem ser a mesma coisa apenas devido à nossa falta de interesse nos casos não-conscientes (ROSENTHAL, 1997, p. 732; ROSENTHAL, 2017, p. 150).

Mas não apenas contextos cotidianos oferecem evidências de que qualidades sensoriais podem existir independentemente de serem conscientes (cf. ROSENTHAL, 1993, p. 357). Também certos fenômenos estudados pelas neurociências e pelas ciências cognitivas apontam nessa direção. Como esclarece Rosenthal:

Na percepção subliminar e visão periférica, e em alguns fenômenos dissociativos como a ‘visão cega’<sup>12</sup>, sujeitos se engajam em padrões de comportamento muito específicos, assim como reportam ocorrências mentais muito específicas. Esses comportamentos e ocorrências mentais são muito próximos daqueles que ocorreriam se os sujeitos tivessem sensações conscientes com certas propriedades qualitativas específicas. Portanto, mesmo que os sujeitos neguem ter tais sensações conscientes, eles se encontram evidentemente em estados que desempenham a maior parte do *papel causal* [grifo meu] que seria desempenhado por certos estados sensoriais conscientes. É natural concluir, portanto, que os estados nos quais esses sujeitos se encontram são, de fato, estados sensoriais que não são conscientes (ROSENTHAL, 1993, p. 357).

Ao mencionar o *papel causal* dos estados mentais que não são sentidos, mas que atuam exatamente como os estados sensoriais conscientes, Rosenthal está sugerindo um critério para a existência de

---

<sup>12</sup> Como explica Claudio Costa: “Trata-se de casos em que uma pessoa, após sofrer lesão na área V1 do córtex visual, torna-se parcialmente cega. Contudo, embora a pessoa afirme nada ver, ela é capaz de adivinhar corretamente muito do que acontece na porção do seu campo visual que perdeu a visão. A explicação seria a de que ela ainda mantém certo grau de consciência perceptual, embora tenha perdido a consciência reflexiva com relação ao campo visual afetado. Ela vê sem saber que vê” (COSTA, 2006, p. 87-8).

fenômenos sensoriais que é *independente da consciência*.<sup>13</sup> Ele entende, ainda, que a ideia de que sensações são invariavelmente conscientes pode ter uma de suas fontes no fato de que nós *classificamos* estados mentais sensoriais e *discernimos* entre as suas várias ocorrências através do modo peculiar “como é” (*what it is like*) para nós estar em tais estados sensoriais, de modo que surge a ideia de que, pelo fato de os classificarmos através desse modo peculiar, não poderiam haver estados sensoriais *desprovidos* desse modo “como é” estar neles.

Mas ele contra argumenta que o simples fato de que classificamos e discernimos esses estados através de seu aspecto consciente – o modo como os sentimos conscientemente – “não mostra que as propriedades por referência às quais nós classificamos os estados sensoriais não podem ocorrer inconscientemente” (ROSENTHAL, 1991, p. 19). Segundo ele, basta comparar o caso de estados sensoriais com o caso de *objetos comuns* captados na percepção, pois nós classificamos e discernimos entre esses objetos por meio do modo como eles aparecem para nós, mas isso não significa que eles não tenham propriedades duradouras, que continuam a existir quando esses objetos não são percebidos. Tomando o exemplo de objetos visíveis, ele explica:

Pode-se dizer que a contribuição que objetos físicos fazem para o modo como eles se parecem para nós é a posse de certas cores, ou mais precisamente, a posse de certos espectros de reflectância. Objetos físicos possuem essas propriedades se alguém os vê ou não (ROSENTHAL, 1991, p. 20)<sup>14</sup>.

Em suma, alegando que meros *interesses pragmáticos* podem ser vistos como a fonte da intuição de que sensações tem de ser estados conscientes, Rosenthal argumenta que nosso uso de certas expressões

---

<sup>13</sup> Veremos a seguir (cf. a nota de rodapé número 21) que, curiosamente, um defensor da essencialidade da consciência para a existência de fenômenos mentais (cf. SEARLE, 1992, p. 18; SEARLE, 1997, p. 31; SEARLE, 1992, p. 84; 1997, p. 125-6) como é John Searle, acaba sugerindo um tal critério (independente da consciência) para a existência de sensações inconscientes.

<sup>14</sup> Já em *Two Concepts of Consciousness* ele havia afirmado, em termos mais gerais: “não temos mais razão para identificar estados mentais com estados conscientes do que temos para identificar objetos físicos com objetos físicos que alguém vê” (ROSENTHAL, 1986, p. 330).

referentes às sensações não implica que elas só possam existir conscientemente, e que fenômenos cotidianos (como sensações intermitentes) assim como fenômenos estranhos estudados pela ciência (como a “visão cega”), nos dão fortes indícios de que as sensações preservam uma existência independente da consciência. O fato de classificarmos os estados sensoriais por meio do modo como os sentimos conscientemente não implica que esses estados sensoriais não preservem propriedades *objetivas*, no sentido de serem propriedades que existem independentemente de serem vivenciadas por um sujeito<sup>15</sup>. Mas, se é assim, que visão ele sustenta a respeito do que são, de fato, as sensações?

### 3. A teoria do espaço de qualidades

Na concepção cartesiana das sensações – isto é, a concepção das sensações que parte do princípio de que a consciência é *essencial* para a existência de todos os fenômenos mentais – prevalece a ideia de que as sensações são conhecidas por nós – ou seja, que nós as sentimos – apenas através da experiência consciente, de uma tal maneira que “não há nada mais para conhecer além do que a consciência revela.” (ROSENTHAL, 2010, p. 373).

Porém, a tal concepção, que ele rotula de teorias da sensação “baseadas na consciência”, ele opõe aquelas que ele chama de “teorias do *papel perceptivo*” [*perceptual-role theories*], teorias estas que não negam que a consciência nos dá, de fato, um acesso às sensações, mas que acentuam que a consciência *não é* a nossa única via de acesso a elas. Como esclarece Rosenthal: “O papel que as qualidades mentais desempenham na percepção nos dá, de acordo com essas teorias, uma rota independente,

---

<sup>15</sup> John Searle caracteriza os estados mentais conscientes como *ontologicamente subjetivos*, no sentido de que eles “somente existem quando são experimentados por algum indivíduo humano ou animal” (SEARLE, 2002, p. 40; Searle, 2010, p. 58). Entendo que, em uma visão como a de Rosenthal, mesmo estados sensoriais são, no fundo, ontologicamente *objetivos*, contrariando a visão de Searle, pois eles preservam propriedades fundamentais que existem independentemente de qualquer vivência consciente. Para um argumento contra a tese da subjetividade ontológica da consciência na filosofia de Searle, cf. Prata (2020).

alternativa, para aprendermos sobre as qualidades mentais” (ROSENTHAL, 2010, p. 374).

Sendo assim, as teorias do papel perceptivo concebem as qualidades mentais – as qualidades que caracterizam nossos estados mentais sensoriais – como sendo “as propriedades em virtude das quais nós fazemos discriminações perceptivas.” (ROSENTHAL, 2010, p. 374). Na medida em que uma *discriminação* – no sentido de um ato de *discernir* entre certas entidades – operada por um sujeito pode ser *constatada* por um observador externo – uma *terceira pessoa* –, através do *comportamento* do sujeito que opera a discriminação, o papel perceptivo das qualidades mentais – ou seja, seu papel nas discriminações perceptivas dos sujeitos percipientes – nos dá um acesso *objetivo* a essas qualidades.

A esse respeito é importante destacar que, como vimos acima (seção 2), a própria percepção é um tipo de fenômeno mental que, frequentemente, ocorre de maneira *inconsciente*, por exemplo na percepção subliminar, em fenômenos dissociativos como a visão cega, ou em certos contextos cotidianos<sup>16</sup>. E se temos bons motivos para aceitar que *percepções* ocorrem inconscientemente, então temos de aceitar também que as *sensações* que compõem tais percepções, com as suas qualidades mentais, também ocorrem de modo inconsciente. Como escreve Rosenthal:

Na verdade, a evidência de pré-ativação [*priming*]<sup>17</sup> em experimentos de laboratório frequentemente mostra que o processamento psicológico inconsciente de um indivíduo distingue cores e outras características qualitativas de estímulos exatamente como na percepção consciente. Há todas

---

<sup>16</sup> Rosenthal menciona a seguinte situação: “Alguém pode estar esperando a chegada de uma pessoa em uma sala lotada e, sem ter notado conscientemente a pessoa, repentinamente dirige o seu olhar diretamente para onde a pessoa está. Tem que ter havido algum estímulo [*input*] visual que permitiu que ele deslocasse o olhar direto para a pessoa e, portanto, [tem que ter havido] qualidades mentais relevantes para a percepção dessa pessoa. Mas uma vez que não se notou antes conscientemente a pessoa, as qualidades mentais ocorreram sem serem conscientes” (ROSENTHAL, 2010, p. 374).

<sup>17</sup> No chamado “efeito de pré-ativação” [*priming effect*] um estímulo é apresentado ao sujeito experimental de um modo rápido demais para que ele seja percebido conscientemente, mas mesmo assim o experimento mostra que esse estímulo é capaz de influenciar processos cognitivos do sujeito posteriores à sua exibição.

as razões para concluir que qualidades mentais ocorrem mesmo quando a percepção falha em ser consciente (ROSENTHAL, 2010, p. 375).

Rosenthal propõe uma forma específica de teoria do papel perceptivo sobre as sensações, a “teoria do espaço de qualidades” (*quality-space theory*) que, segundo ele, é capaz de classificar as qualidades mentais de um maneira (pelo menos) tão acurada quanto as teorias baseadas na consciência, por mais que tal teoria do espaço de qualidades não possa explicar porque, algumas vezes, há um modo “como é” vivenciar as qualidades mentais. Mas, para explicar isso, Rosenthal já tem a sua teoria dos pensamentos de ordem superior<sup>18</sup>.

A teoria do espaço de qualidades parte da distinção fundamental entre (1) as propriedades perceptíveis dos objetos – propriedades físicas que existem independentemente de nós – e (2) as qualidades mentais que caracterizam nossos estados mentais *sensoriais*. Por um lado, temos propriedades como cores, tamanho visível, forma e localização (perceptíveis pela *visão*), assim como propriedades como tons, timbres e origens dos sons (perceptíveis pela *audição*), e outras propriedades perceptíveis (através de outras modalidades sensoriais). Por outro lado, temos nossas qualidades mentais, correspondentes a tais propriedades perceptíveis. A correspondência entre os dois diferentes tipos de propriedades é decisiva para que o sujeito percipiente seja capaz de realizar discriminações perceptivas. Como explica Rosenthal:

Para discriminar entre propriedades perceptíveis uma criatura tem que ser capaz de se encontrar em estados que diferem ao menos tão finamente quanto as propriedades perceptíveis sendo discriminadas. E as diferenças relativas entre aqueles estados tem que refletir as diferenças entre as propriedades perceptíveis (ROSENTHAL, 2010, p. 377).

---

<sup>18</sup> Conforme foi dito na introdução do presente trabalho, por motivos de direcionamento temático, não abordarei a maneira como Rosenthal pretende explicar o caráter fenomenal das qualidades mentais através da sua teoria dos pensamentos de ordem superior. Discutirei esse tema em trabalhos que estão em preparação.

Se houver tal correspondência entre as propriedades dos dois domínios, então seremos capazes de determinar, a partir de um ponto de vista de terceira pessoa – ou seja, *objetivo* – se a criatura é capaz de discernir entre essas propriedades que ela é capaz de perceber. Assim, de acordo com Rosenthal, “isso nos permite construir um *espaço de qualidades* [grifo meu] no qual a distância entre duas propriedades perceptíveis quaisquer é uma função de quantas propriedades entre as duas a criatura pode discernir” (ROSENTHAL, 2010, p. 377), sendo que teremos diferentes espaços de qualidades correspondentes às diferentes modalidades sensoriais (p. ex., visão, audição, olfato, etc.) que a criatura possui.

É importante destacar que a teoria do espaço de qualidades diz respeito não apenas a *sensações perceptivas* – que compõem as percepções de objetos e estados de coisas no ambiente externo – mas também as *sensações corporais*, como dores, cócegas ou coceiras (cf. ROSENTHAL, 1986, p. 348), já que nós também somos capazes de discernir entre elas, o que nos permite elaborar teoricamente um espaço de qualidades (corporais) que corresponderia a essas discriminações.

E, de acordo com Rosenthal, uma diferença importante entre as teorias das sensações baseadas na consciência e as teorias do papel perceptivo é que estas últimas *não comparam* o espaço de qualidades – de cada modalidade sensorial – com as propriedades perceptíveis externas, mas sim *extrapolam* a partir do espaço de qualidades das propriedades perceptíveis para determinar o espaço das qualidades *mentais* correspondentes, pois a teoria do espaço de qualidades rejeita a alegação – fundante das teorias baseadas na consciência – de que só temos um acesso de primeira pessoa às qualidades mentais. É por isso que a teoria do espaço de qualidades aborda essas qualidades mentais pela via das *discriminações* perceptivas de que o sujeito é capaz e, a partir disso, reconstitui teoricamente quais seriam as qualidades correspondentes a essas discriminações. Portanto, nas palavras de Rosenthal: “A teoria do espaço de qualidades classifica [*taxonomizes*] as qualidades mentais

através de seu papel em tais discriminações” (ROSENTHAL, 2010, p. 379), o que, segundo ele, permite a essa teoria classificar essas qualidades mentais com uma acurácia maior do que qualquer teoria baseada na consciência. E essa superioridade da teoria proposta por Rosenthal teria sérias consequências para os argumentos de conceitabilidade contra o fisicalismo, pois, na visão dele:

Uma teoria do papel perceptivo, [...], acomoda a ocorrência de qualidades mentais em percepções inconscientes, e desse modo enfraquece qualquer abordagem baseada na consciência. Assim nós podemos esperar que nenhuma teoria do papel perceptivo, tal como a teoria do espaço de qualidades, vá acomodar a inversão não detectável. E a teoria do espaço de qualidades dá um passo além ao dar *razões concretas* [grifo meu] pelas quais tal inversão é impossível (ROSENTHAL, 2010, p. 380).

Rosenthal enfatiza que a ideia de que uma *inversão não detectável* de qualidades mentais é algo no mínimo, *concebível* é tida como óbvia por muitos estudiosos em filosofia da mente. Mas ele entende que considerar a conceitabilidade da inversão de qualidades como “óbvia” é uma consequência da aceitação tácita da ideia de que “a consciência tem a primeira e a última palavra a respeito das qualidades mentais” (ROSENTHAL, 2010, p. 370), ou seja, é uma consequência da pressuposição de que “qualidades mentais só permitem um acesso de primeira pessoa” (ROSENTHAL, 2010, p. 369), um acesso por meio do modo peculiar “como é” (*what it is like*) vivenciar essas qualidades do nosso próprio ponto de vista de primeira pessoa.

Em sua teorização sobre as qualidades mentais, a tentativa dele é mostrar que um outro acesso aos estados sensoriais é possível, um acesso independentemente do modo como, usualmente, os conhecemos em nossas experiências conscientes. E, em seguida, ele procurará mostrar como tal acesso *objetivo*, independente da consciência (que é subjetiva), desfere um golpe fatal na ideia de que uma inversão *indetectável* de

qualidades mentais é *concebível*, pois toda inversão seria, pelo contrário, *detectável*.

E, de acordo com Rosenthal, a viabilidade de um acesso objetivo, de terceira pessoa, às qualidades mentais também tem graves consequências para outros argumentos baseados na *conceptibilidade* da desconexão entre o mental e o físico, como o argumento dos zumbis, de David Chalmers (1996), e sua ideia de um “problema difícil” a respeito da consciência fenomenal (cf. CHALMERS, 1995; CHALMERS, 2020), bem como a “lacuna explanatória” de Joseph Levine (1983), e o “argumento do conhecimento” de Frank Jackson (1982) (cf. ROSENTHAL, 2010, p. 370-72)<sup>19</sup>.

Porém, não será possível, por motivos de espaço, discutir aqui esse tema. Para concluir o presente trabalho, irei expor uma rápida avaliação dos argumentos oferecidos por Rosenthal para fundamentar a crença na existência de sensações inconscientes – crença decisiva para o sucesso de sua teoria da consciência, a “teoria dos pensamentos de ordem superior”.

#### 4. Considerações finais

Já em seu conhecido artigo *Two Concepts of Consciousness*, Rosenthal havia alertado que “a consciência parece central para a mentalidade somente porque ela é tão básica para o modo como conhecemos nossos estados mentais. Mas o modo como conhecemos as coisas frequentemente não é um guia confiável para a sua natureza.” (ROSENTHAL, 1986, p. 330).

---

<sup>19</sup> Rosenthal (2010, p. 371) menciona ainda o famoso “argumento do moinho”, exposto no parágrafo 17 da *Monadologia* de Gottfried Leibniz. Leibniz argumentou que nem a percepção nem nada que dependa dela pode jamais ser explicado em termos mecânicos, pois se entrássemos em uma máquina, do tamanho de um moinho, pretensamente capaz de gerar pensamentos e percepções, não veríamos dentro dela nada que pudesse explicar esses fenômenos mentais, pois veríamos apenas peças empurrando umas as outras. Porém, na visão de Rosenthal: “A alegação de Leibniz de que não veríamos nada dentro da máquina que explicasse a percepção simplesmente captura de uma maneira especialmente memorável a visão de que as qualidades mentais que ocorrem na percepção são acessíveis apenas através da consciência” (ROSENTHAL, 2010, p. 371). Para uma pequena e interessante discussão do argumento de Leibniz, cf. Searle (1983, p. 267-8) e Searle (1995, p. 371-2).

No meu modo de entender, o reconhecimento de que (i) a experiência consciente é o que nos dá o acesso mais imediato a nossos próprios estados mentais, somado aos numerosos indícios de que (ii) tais estados efetivamente *existem* de maneira inconsciente, realmente desferiu um duro golpe na visão cartesiana discutida ao longo do presente trabalho, pois a existência de fenômenos mentais inconscientes nos mostra com clareza que o acesso imediato fornecido pela consciência – por mais que seja o que há de mais familiar a nós mesmos (cf. CHALMERS, 1996, p. 3) – é *falho*.

Se é verdade, por exemplo, que efetivamente *vejo* a parede branca diante de mim neste momento, e se é também verdade que essa visão me dá acesso a algumas características desse objeto externo, por outro lado, não se pode negar que a parede certamente possui inúmeras outras características – como o exato *peso* dessa porção de parede que eu vejo, a exata *composição dos tijolos*, ou o *tipo de massa* usado para fixá-los – que escapam à minha visão. Os indícios de que estados mentais – inclusive sensações – existem sem que os vivenciemos conscientemente (cf. a seção 2, acima) indicam que o conhecimento acerca dos nossos fenômenos mentais oferecido pela consciência não abarca completamente o que tais fenômenos são, abrindo espaço para a possibilidade de que mesmo as sensações podem existir independentemente de qualquer experiência consciente.

É possível, é claro, tentar negar que nossos próprios estados mentais sensoriais possam ser comparados a um objeto externo, como uma parede. É possível argumentar que nossos estados sensoriais, no que são (a) caracterizados por um *aspecto qualitativo* – que parece ser necessariamente consciente (cf. a seção 1, acima) – aspecto este que (b) nos aparece de modo *imediato* – sem intermediários<sup>20</sup> – são estados

---

<sup>20</sup> Argumentando que os teóricos da identidade entre o mental e o físico *não podem* descartar a contingência da identidade entre fenômenos mentais e fenômenos físicos tal como eles podem descartar essa contingência no caso da identificação entre movimentos moleculares e calor, Kripke afirma: “No caso do movimento molecular e do calor, há algo, a saber, a sensação de calor, que é um intermediário entre o fenômeno externo e o observador. No caso do mental e do físico, *não é possível haver um intermediário* [grifo meu], já que aqui o fenômeno físico é suposto como idêntico ao próprio fenômeno interno” (KRIPKE 2001 [1972], p. 151-2).

*indissociáveis* do modo peculiar como nós os vivenciamos – ao passo que a parede é obviamente dissociável dos estados mentais através dos quais nós a conhecemos.

Entretanto, as circunstâncias nas quais evidências *objetivamente* constatáveis – como os *comportamentos* que nos levam a concluir a presença de *percepção subliminar*, *dores inconscientes*<sup>21</sup> ou o curioso fenômeno da *visão cega* – indicam a presença de sensações e percepções inconscientes, são circunstâncias que mostram que as sensações possuem, sim, uma existência anterior a qualquer vivência subjetiva, sendo, portanto, dissociáveis da maneira como nós os vivenciamos.

Quando Rosenthal argumenta que nossa *intuição* (isto é, nossa *compreensão pré-teórica*) de que sensações só existem em forma consciente é uma consequência de nossos *interesses pragmáticos*, isso me parece razoável, já que tais interesses, e nossas *estratégias de ação* para alcançá-los, certamente desempenham um papel em nossas formas espontâneas de compreender o mundo. Se sensações desprovidas de consciência não desempenham nenhum papel em nossa avaliação das condições de nosso próprio corpo ou de outros, parece razoável pensar que nada em nossa vida cotidiana nos levará a pensar na existência de tais sensações inconscientes. Apenas considerações teóricas, como as expostas por Rosenthal, é que poderiam nos levar a considerar a existência de sensações das quais não temos consciência.

Além disso, quando Rosenthal argumenta que nosso uso de certas expressões, designadoras de estados sensoriais, que parecem implicar

---

<sup>21</sup> É curioso notar que mesmo um defensor da visão cartesiana da consciência – como essencial para a existência de fenômenos mentais – tal como John Searle, defende a existência de sensações inconscientes (cf. SEARLE, 1992, p. 164-7; SEARLE, 1997, p. 236-40), o que me parece ser uma incoerência em sua filosofia da mente. Ele menciona o seguinte exemplo: “Certa vez tive uma fratura no pulso. Esse ferimento causava muita dor durante o dia e a dor aumentava se eu não fosse cuidadoso quanto os movimentos que eu fazia com o braço. Eu percebi uma coisa interessante ao dormir. Eu estava totalmente adormecido, de modo que eu não sentia dor nenhuma, e ainda assim meu corpo se movia durante o sono de modo a proteger o ferimento. Como deveríamos descrever tal caso? [...]. Nós normalmente não falamos sobre dores inconscientes, mas facilmente poderíamos, e casos como esses nos dariam a motivação para falar de dores inconscientes” (SEARLE, 2004, p. 244). É interessante notar que, ao considerar justificado falar de “dores inconscientes”, Searle está aceitando como critério da presença de um fenômeno mental sensorial o *papel causal* desempenhado por esse fenômeno no comportamento do sujeito, independentemente de qualquer vivência subjetiva. Sobre a concepção de Searle a respeito do inconsciente, cf. Prata (2017c), Prata (2017d) e Prata (2019).

consciência desses estados, *não mostra* que os estados sensoriais que elas designam não podem existir inconscientemente, isso me parece convincente, afinal nosso *uso* de quaisquer expressões linguísticas se insere em nossas ações no mundo que nos cerca, de modo que é de se esperar que esse uso não alcance fenômenos (inconscientes) que não se apresentam a nós, e que só podem ser conhecidos através de sofisticadas construções teóricas.

Em outros termos, se nosso uso coloquial de certas expressões, como “sensação” ou “doloroso”, só ocorre em contextos concretos de ação, e se fenômenos inconscientes não se manifestam diretamente a nós em tais contextos, exigindo elaborações teóricas para que possamos *inferir* sua existência, é de se esperar que nosso uso cotidiano não alcance tais fenômenos inconscientes.

De fato, parece perfeitamente razoável considerar que quando há, por exemplo, um ferimento doloroso, mas alguma distração nos faz “esquecer” da dor por certo tempo, essa dor preserva uma existência *independente da consciência*, até que nossa atenção se volte para ela novamente. E o mesmo tipo de existência independente da consciência parece caracterizar os fenômenos sensoriais que podemos inferir de situações como aquelas que sugerem haver *percepção subliminar* ou a bizarra *visão cega*, já que evidências comportamentais nos levam a inferir que o sujeito que se encontra em tais situações de fato tem percepções, ainda que ele não as vivencie.

É bem verdade que, para que a perspectiva de Rosenthal a respeito das sensações se mostre realmente sustentável, é necessário verificar se a sua teoria dos pensamentos de ordem superior é realmente capaz de explicar o aspecto fenomenal das experiências sensoriais. Mas essa verificação é uma tarefa que será deixada para oportunidades futuras.

## Referências

ADAM, Ch. & TANNERY, P. (Orgs.). **Ouvres de Descartes – Meditations de prima philosophia**. Paris: Vrin, 1996a. (Vol. VII).

\_\_\_\_\_. (Orgs.). **Ouvres de Descartes – Méditations et principes** (traduction française). Paris: Vrin, 1996b. (Vol. IX).

ARMSTRONG, A. What is Consciousness? In: BLOCK, N.; FLANAGAN, O. & GÜZELDERE, G. (Orgs.). **The Nature of Consciousness: philosophical debates**. Cambridge (Massachusetts): MIT Press, 1997, p. 721-8.

\_\_\_\_\_. **The mind-body problem: An opinionated introduction**. Boulder (CO); Oxford (UK): Westview Press, 1999.

ATKINSON, R. L.; ATKINSON, R. C.; SMITH, E. E.; BEM, D. & NOLEN-HOEKSEMA, S. **Introdução à psicologia de Hilgard**. 13.ed. Porto Alegre: Artmed, 2002.

\_\_\_\_\_. (Eds.). **Atkinson and Hilgard's introduction to psychology**. 15th Edition. Andover (UK): Cengage Learning EMEA, 2009.

BARGH, J. A.; MORSELLA, E. The Unconscious Mind. In: **Perspectives on Psychological Science**, Vol. 3, No 1, pp. 73-79, 2008.

BOAG, S. In defense of unconscious mentality. In: BOAG, S.; BRAKEL, L. A. W. & TALVITIE, V. (Orgs.). **Psychoanalysis and philosophy of mind: unconscious mentality in the twenty-first century**. London: Karnac Books, 2015, p. 239-65.

BRENTANO, F. **Psychologie vom empirischen Standpunkt**. 2.ed. Leipzig: Meiner, 1924 [original: 1874].

\_\_\_\_\_. **Psychology from an empirical standpoint**. London: Routledge & Kegan Paul, 1995.

CHALMERS, D. Facing up to the problem of consciousness. **Journal of Consciousness Studies**, Vol. 2, Nº 3, p. 200-19, 1995.

\_\_\_\_\_. **The conscious mind: in search of a fundamental theory**. Oxford: Oxford University Press, 1996.

\_\_\_\_\_. Enfrentando o problema da consciência. **Prometeus**, Nº 33, maio-agosto, p. 319-52, 2020.

COSTA, C. F. Definindo consciência. **Princípios**, Vol. 13, Nº. 19-20, p. 81-101, jan./dez., 2006.

COTTINGHAM, J; STOOOTHOF, R. & MURDOCH, D. **The philosophical writings of Descartes**. Cambridge: Cambridge University Press, 1984. (Vol. II).

DESCARTES, R. **Discurso do método; Meditações; Objeções e respostas; As paixões da alma; Cartas**. 2ª ed. São Paulo: Abril Cultural, 1979. (Os Pensadores).

GENNARO, R. **Consciousness and selfconsciousness: a defense of the higher order thought theory of consciousness**. Amsterdam/Philadelphia: John Benjamins Publishing, 1996.

JACKSON, F. Epiphenomenal qualia. **Philosophical Quarterly**, 32, p. 127-36, 1982.

KRIPKE, S. **Naming and necessity**. Cambridge (Mass.): Harvard University Press, 2001 [original: 1972].

LEVINE, J. Materialism and qualia: the explanatory gap. **Pacific Philosophical Quarterly**, 64, p. 354-61, 1983.

MARQUES, E. Consciência e autoconsciência em Leibniz. **Analytica** (UFRJ), Vol. 21 nº 1, p. 67-83, 2017.

MASLIN, K. T. **An introduction to the philosophy of mind**. Cambridge (UK): Polity Press, 2001.

\_\_\_\_\_. **Filosofia da mente**. Porto Alegre: Artmed, 2009.

PRATA, T. A. A problemática da subjetividade e a ontologia da consciência. In: AMORA, K. C.; BRILHANTE, A. A. & COSTESKI, E. (Eds.). **Extratos filosóficos: 10 anos do curso de pós-graduação em filosofia da UFC**. Fortaleza: Edições UFC, 2009, p. 45-58.

PRATA, T. A. Um argumento contra a tese da subjetividade ontológica da consciência no naturalismo biológico de John Searle. In: **Filosofia Unisinos**. Vol. 21, No 3, pp. 303-11, sep/dec, 2020.

\_\_\_\_\_. Saul Kripke sobre o argumento de Descartes para o dualismo mente-corpo. **Cadernos UFS de Filosofia**, Vol. 10, p. 17-29, 2011.

- \_\_\_\_\_. A estrutura da autoconsciência na filosofia da mente de John Searle. **Veritas** (PUCRS), Vol. 62, Nº 2, maio-ago., p. 428-52, 2017a.
- \_\_\_\_\_. Apresentação do artigo “Uma teoria da consciência” de David Rosenthal. **Perspectiva Filosófica**, Vol. 44, Nº 2, p. 133-42, 2017b.
- \_\_\_\_\_. Uma crítica à concepção disposicional de Searle sobre os fenômenos mentais inconscientes. In: Araújo, A. [et al.] (Orgs.). **Pragmatismo, filosofia da mente e filosofia da neurociência**. São Paulo: ANPOF, p. 387-403, 2017c.
- \_\_\_\_\_. A concepção disposicional do inconsciente na filosofia da mente de John Searle. **Reflexões** (UVA), Ano 6, Nº 11, Julho a Dezembro, p. 201-16, 2017d.
- |\_\_\_\_\_. Uma objeção à concepção disposicional dos fenômenos mentais inconscientes. **Principia** (UFSC), Vol. 22, Nº 3, p. 507-26, 2018.
- \_\_\_\_\_. A teoria disposicional de Searle sobre os fenômenos inconscientes e o problema da eficácia causal. **Pensando – Revista de Filosofia** (UFPI), Vol 10 Nº 19, p. 11-25, 2019.
- ROSENTHAL, D. Two concepts of consciousness. **Philosophical Studies**, 49, p. 329-59, 1986.
- \_\_\_\_\_. The independence of consciousness and sensory quality. **Philosophical Issues**, 1, p. 15-36, 1991.
- \_\_\_\_\_. State consciousness and transitive consciousness. **Consciousness and Cognition**, 2, p. 355-63, 1993.
- \_\_\_\_\_. A theory of consciousness. In: BLOCK, N.; FLANAGAN, O. & GÜZELDERE, G. (Orgs.). **The nature of consciousness: philosophical debates**. Cambridge (Massachusetts): MIT Press, 1997, p. 729-53, 1997.
- \_\_\_\_\_. Consciousness and its function. **Neuropsychologia**, 46, p. 829-40, 2008.
- \_\_\_\_\_. How to think about mental qualities. **Philosophical Issues**, Vol. 20, p. 368-93, 2010.
- \_\_\_\_\_. Uma teoria da consciência. **Perspectiva filosófica**, Vol. 44, Nº 2, p. 143-78, 2017.

SEARLE, J. R. **Intentionality**: An essay in the philosophy of mind. Cambridge: Cambridge University Press, 1983.

\_\_\_\_\_. Consciousness, unconsciousness, and intentionality. **Philosophical Topics**, Vol. XVII, p. 193-209, 1989.

\_\_\_\_\_. **The rediscovery of the mind**. Cambridge Mass., London: MIT Press, 1992.

\_\_\_\_\_. **Intencionalidade**. São Paulo: Martins Fontes, 1995.

\_\_\_\_\_. **A redescoberta da mente**. São Paulo: Martins Fontes, 1997.

\_\_\_\_\_. **Consciousness and language**. Cambridge (UK): Cambridge University Press, 2002.

\_\_\_\_\_. **Mind**: A brief introduction. Oxford: Oxford University Press, 2004.

\_\_\_\_\_. **Consciência e linguagem**. São Paulo: Martins Fontes, 2010.

\_\_\_\_\_. **Seeing things as they are**: a theory of perception. Oxford: Oxford University Press, 2015.

TALVITIE, V. **Freudian unconscious and cognitive neuroscience**: from unconscious fantasies to neural algorithms. London: Karnac Books, 2009.

TYE, M. Philosophical problems of consciousness. In: VELMANS, M. & SCHNEIDER, S. (Orgs.). **The Blackwell companion to consciousness**. Oxford: Blackwell, 2007, p. 23-35.



## Representacionalismo e realismo direto na teoria da percepção de John R. Searle

*João Paulo M. de Araujo*<sup>1</sup>

A teoria da percepção de Searle possui ao menos duas versões. A primeira é oriunda dos anos oitenta e está presente no segundo capítulo de seu livro *Intencionalidade* (1983). A segunda é bem recente, sendo, portanto, tema exclusivo de investigação em seu livro *Seeing Things As They Are* (2015). De acordo com Searle, sua atual versão pode ser entendida como uma continuação da linha de análise que ele começou no capítulo 2 de *Intencionalidade*. Nessa atual versão, que podemos aqui considerar ampliada e revisada, Searle (2015, p. 5) teria por tarefa acabar com algumas obscuridades e incompletudes de sua explicação em *Intencionalidade*. Na explicação de Searle sobre a percepção, existe uma relação intrínseca entre intencionalidade e percepção de tal modo que o fenômeno perceptual não pode ser concebido senão como intencional. Em sua teoria, Searle defende um realismo direto. O realismo direto é uma teoria filosófica que afirma que o tipo de acesso que nós temos aos objetos do mundo externo ocorre de forma direta e imediata. Uma visão filosófica

---

<sup>1</sup> Doutor em filosofia pelo Programa de Doutorado Interinstitucional UFPB-UFPE-UFRN e professor substituto na Universidade Federal de Pernambuco (UFPE).

contrária ao realismo direto é o realismo indireto ou teoria representativa da percepção. Essa teoria afirma que percebemos indiretamente objetos físicos no mundo, ao passo que percebermos imediata e diretamente objetos internos não físicos, como *sense-data*, ideias, impressões etc. (MAUND, 2003). A teoria da percepção de Searle, em ambas as versões, pode ser entendida como a tese de que a percepção é intencional e que o tipo de acesso que temos aos objetos do mundo real ocorre de forma direta, por meio de uma relação causal intencional. Searle defende também que existe um elemento fenomenal comum entre a percepção de um objeto real no mundo e uma alucinação. Portanto, este capítulo consiste em dois movimentos, primeiro uma resumida e circunscrita exposição de sua atual teoria da percepção levando em consideração aquilo que em linhas gerais considero seus principais aspectos, a saber: (1) uma defesa do realismo direto, (2) a tese de que a percepção é intencional e (3) a tese do elemento comum. E em segundo, uma crítica ao seu modelo explicativo onde apresentarei algumas considerações que apontam para uma inconsistência entre (1) e (3) que é coextensiva à (1) e (2) devido a uma relação intrínseca entre (2) e (3).

Em minha leitura de Searle, parto da compreensão de que a intencionalidade é um elemento que está presente em todos os aspectos do seu pensamento filosófico. Por conseguinte, no que concerne à sua teoria da percepção, o meu problema objetiva responder a seguinte pergunta: é consistente a explicação intencional da percepção [(2) e (3)] em Searle com a sua defesa de um realismo direto (1)? Esta pergunta pode ser formulada tanto na primeira versão de sua teoria quanto na segunda. Nas duas versões, Searle defende que existe um elemento comum entre o caráter fenomenal de uma percepção verídica e o caráter fenomenal de uma alucinação. Num primeiro exame, este ponto pode não parecer problemático, contudo, se levarmos algumas discussões contemporâneas sobre a percepção a sério, perceberemos que é inviável defender um realismo direto e sustentar a tese do elemento comum no mesmo bloco explicativo, para isto, basta uma breve pesquisa na literatura disjuntivista

sobre a percepção. Em linhas gerais, um disjuntivista afirma que não existe um elemento comum entre uma percepção verídica e uma alucinação. Para um disjuntivista, é inconsistente sustentar uma defesa do realismo direto e endossar a tese do elemento comum. Por esta razão, disjuntivistas não endossam a tese do elemento comum, mas defendem um realismo direto. Em minha interpretação, argumentarei que apesar da teoria de Searle possuir um forte apelo intuitivo e explicativo acerca da tese de que a percepção é intencional e direta, ela pode ser interpretada como uma teoria representativa devido às inconsistências das relações entre (1), (2) e (3).

### **O realismo direto e a intencionalidade da percepção**

No que concerne aos problemas difíceis de tratar em filosofia da mente, o problema da intencionalidade só está abaixo do problema da consciência, sendo o próprio problema da intencionalidade considerado por Searle como “uma espécie de reflexo do problema da consciência” (SEARLE, 2004, p. 159). A intencionalidade<sup>2</sup> é definida por Searle como uma “propriedade de muitos estados e eventos mentais pela qual eles são direcionados para, ou acerca de, objetos e estados de coisas no mundo” (SEARLE, 1983, p. 1). Por outro lado, existe a *percepção* que, do ponto de vista pragmático, desempenha um enorme papel no nosso cotidiano. Agimos com base em crenças, desejos, intenções etc. No entanto, tudo isso só é possível porque possuímos uma relação perceptual com o mundo à nossa volta ou, para sermos filosoficamente precisos, com o mundo externo, como geralmente é caracterizado pelos filósofos. Portanto, classicamente, o problema da percepção está relacionado à maneira como

---

<sup>2</sup> Vale ressaltar que Searle apesar de seguir uma tradição que modernamente remonta à Brentano, sua explicação da intencionalidade se distancia em ao menos três aspectos dessa tradição: (1) Apenas alguns estados mentais (e não todos) têm intencionalidade; (2) Intencionalidade não é a mesma coisa que consciência; (3) O pretender e as intenções são apenas uma forma de Intencionalidade entre outras e não possuem uma *status* especial. Para um desdobramento dos três aspectos Cf. Searle, 1983, p. 1-3.

nossas experiências perceptuais se relacionam com os objetos do mundo externo.

Em um *paper* que antecede a publicação de seu recente livro sobre a percepção, Searle (2012, p. 9) afirma que “a filosofia geralmente está preocupada com a relação dos seres humanos com a realidade, e o meio consciente primário pelo qual os seres humanos se relacionam com o mundo real é o da experiência perceptual.” Em *Seeing Things As They Are* (2015), Searle apresenta o problema da percepção através de uma discussão com alguns filósofos modernos. A estratégia de Searle consiste em apresentar esses filósofos como endossando um tipo de visão que seria insuficiente para dar conta do problema da percepção. Sua teoria da Intencionalidade Perceptual (*Perceptual Intentionality*) seria uma nova perspectiva que teria por tarefa explicar como a percepção funciona.

Nessa nova perspectiva, Searle irá explicitamente defender um realismo direto. Por outro lado, durante todo o livro *Intencionalidade*, o único momento em que Searle tem a necessidade de afirmar explicitamente que ele defende um realismo direto é na seção IV do capítulo dois. Nessa seção, Searle (1983, p. 58) faz uma ‘digressão’ do problema da percepção na tradição filosófica apenas com o intuito de comparar sua visão com aquilo que ele chama de rivais históricas de sua explicação da percepção. Além do mais, Searle (1983, p. 57) é resolutivo ao afirmar que “a explicação da percepção visual que tenho discutido até agora é, *acho*, uma versão do realismo “ingênuo” [direto de senso comum].”

É comum observarmos em seus escritos alguns rótulos para expressar “errôneas” visões acerca de determinadas posturas filosóficas. Em sua obra *A Redescoberta da Mente* (1992), Searle lançou mão de uma antagonista postura batizando-a de *dualismo conceitual*, a saber, uma série de equívocos cometidos pelos filósofos, ao longo dos anos, ao discutirem a natureza da consciência e a relação entre mente e cérebro. Segundo Searle (1992, p. 26), essa visão (que remonta a Descartes) consistia na ideia de que em algum sentido importante, “físico” implica em algo “não-mental”

e "mental" implica em algo "não-físico" e dessa maneira poderíamos aplicar essa concepção tanto ao dualismo quanto ao materialismo. Analogamente à noção de dualismo conceitual, em *Seeing Things As They Are*, Searle (2015) apresenta uma série de considerações acerca do que ele chamou de *Bad Argument*. No que consiste o *Bad Argument*? *Prima facie*, podemos caracterizá-lo como uma errônea postura (acerca da percepção) sustentada por muitos filósofos ao longo dos séculos. O *Bad Argument* pode ser compreendido a partir da ideia de que “nós nunca percebemos diretamente objetos e estados de coisas no mundo, mas percebemos apenas diretamente nossas experiências subjetivas.” (SEARLE, 2015, p. 11). Por essa razão, Searle reserva ao *Bad Argument* o lugar de um dos maiores erros cultivado pelos filósofos ao longo dos séculos sobre o problema da percepção. Assim, como no caso do *dualismo conceitual*, temos também no *Bad Argument* Descartes como um dos principais expoentes deste grande erro. No que concerne ao *Bad Argument*, temos a seguinte passagem:

No que se segue, pela expressão "*Bad Argument*", eu tenho em mente qualquer argumento que tente tratar a experiência perceptiva como um objeto de experiência real ou possível. E também usarei a expressão para nomear a conclusão do argumento, que nunca vemos objetos materiais diretamente (SEARLE, 2015, p. 29).

Mas o que significa dizer que não percebemos diretamente os objetos e estados de coisas no mundo, mas apenas nossas experiências subjetivas? Parece razoável sustentar que nosso acesso ao mundo objetivo ocorre indiretamente e que tudo o que filtramos da realidade não passa de representações que estão relacionadas ao modo como percebemos as coisas, devido à nossa constituição cognitiva. Contudo, é justamente essa visão que Searle pretende erradicar com sua teoria da percepção, mas para isso, uma série de protocolos devem ser cumpridos, uma vez que sua teoria dispõe de um grande arcabouço conceitual para explicar minuciosamente todo o itinerário filosófico do fenômeno da percepção visual. Esse arcabouço conceitual preserva muitas explicações que já estavam

presentes em *Intencionalidade*, porém, Searle recorre a muitos outros conceitos advindos de outras fases de seu pensamento, como a ontologia social, racionalidade etc.

Para introduzir o ponto de vista searleano, vamos pensar na seguinte situação: suponha que estamos em um ambiente com boas condições de luz, no qual podemos discriminar adequadamente cada objeto percebido. Se, por exemplo, estivermos numa sala de estar, podemos ver objetos comuns a esse tipo de ambiente, tais como mesa, cadeiras, sofá, tapete, livros etc. Poderíamos descrever esses objetos com suas propriedades, além de alguns fatos óbvios acerca do estado de coisas desses objetos, suas relações etc. Em termos perceptuais de senso comum, é razoável supor que estamos percebendo de maneira direta esses objetos e estados de coisas; também, parece razoável supor que esses mesmos objetos e estados de coisas possuem uma existência independente da percepção que temos deles. Em outras palavras, ao sairmos desse ambiente, os objetos não deixarão de existir, pois “os objetos e estados de coisas têm uma existência independente, no sentido que eles existem independentemente de serem experienciados por nós” (SEARLE, 2015, p. 12). O fato mais óbvio sobre isto é que se decidirmos fechar os nossos olhos, a única coisa que irá cessar será a nossa percepção visual desses objetos, e não a existência dos mesmos. Nesse sentido, duas noções são imprescindíveis para entender a relação entre existência e percepção. Searle (2015) chama de ontologicamente objetivo (*ontologically objective*) os estados de coisas que percebemos diretamente no mundo que possuem uma existência que independe de nossa percepção, e de ontologicamente subjetivo (*ontologically subjective*) as experiências que temos desses objetos e estados de coisas. Vale ressaltar que esses dois conceitos fazem parte das considerações de Searle sobre a filosofia da mente e a construção da realidade social. Agora eles desempenham uma função importante na construção de uma teoria da percepção:

Algumas entidades – montanhas, moléculas e placas tectônicas, por exemplo – possuem uma existência independente de qualquer experiência. Elas são ontologicamente objetivas. Mas outras - dores, cêcegas e coceiras, por exemplo - existem apenas na medida em que são experimentadas por um sujeito humano ou animal. Elas são ontologicamente subjetivas (SEARLE, 2015, p. 16).

Todavia, para descrições de percepções ocorre algo que não seria domínio exclusivo da esfera subjetiva, uma vez que, na relação perceptual, a experiência visual subjetiva dependeria dos objetos do mundo objetivo. Searle (2015) chamou atenção para aquilo que os filósofos denominaram de *Argumento da Transparência* (*Argument from Transparency*). Um truísmo importante é que se tentarmos descrever a realidade objetiva juntamente com seus estados de coisas e depois tentarmos descrever nossa experiência subjetiva correspondente àquilo que estamos vendo, não teríamos nenhuma surpresa, uma vez que as duas descrições serão praticamente as mesmas. Eu poderia, por exemplo, descrever uma cena objetiva dizendo “há um notebook preto sobre a mesa branca” e, por outro lado, eu também poderia descrever minha experiência visual subjetiva dizendo “eu vejo um notebook preto sobre a mesa branca.” “A experiência subjetiva tem um conteúdo, o que os filósofos chamam de conteúdo intencional, e a especificação do conteúdo intencional é o mesmo que a descrição do estado de coisas que o conteúdo intencional apresenta” (SEARLE, 2015, p. 12-3). Ainda segundo Searle (2015, p. 59), o argumento da transparência foi enganosamente usado como um argumento contra a intencionalidade da percepção, porém, esse argumento se tornou um forte aliado a favor de sua explicação.

Em outras palavras, a especificação da minha crença de que está fazendo mais de 25 °C de temperatura em Recife, usa, de acordo com Searle, as mesmas palavras na mesma ordem da descrição do fato no mundo que minha crença representa, a saber: está fazendo mais de 25 °C de temperatura em Recife. O conteúdo intencional da experiência subjetiva tem um estado de coisas correspondente no mundo, isto é, uma relação entre percepção e objeto que qualquer pessoa em condições semelhantes

de percepção descreveria, quase que com as mesmas palavras, o que estaríamos percebendo.

Quando a visão está fazendo corretamente seu trabalho biológico, a descrição do conteúdo intencional e a descrição do estado de coisas que é apresentado deve ser a mesma, porque uma crucial [main] função biológica da experiência perceptiva é fornecer a você conhecimento sobre o mundo real (SEARLE, 2015, p. 13).

Na passagem acima temos algo que a essa altura já está claro, isto é, a postura naturalista de Searle. A percepção (assim como a intencionalidade) está ancorada numa condição biológica, que, de um ponto de vista evolutivo, tratou de estabelecer um vínculo relacional entre o fenômeno da experiência visual e o mundo com seus inúmeros objetos. Portanto, nesse jogo evolutivo de objetos e de conteúdos da percepção, temos uma relação direta com a realidade. O *Bad Argument*, como foi descrito acima, seria uma errônea maneira de descrever nossas relações perceptuais com o mundo.

Consequentemente, qualquer filosofia que tenha pressupostos naturalistas possui um comprometimento direto ou indireto com alguma forma de realismo. Vimos que na definição de Searle (2015, p. 11), o que caracteriza o *Bad Argument* é justamente a ideia de que nós nunca percebemos diretamente objetos e estados de coisas no mundo, percebemos apenas nossas experiências subjetivas. Ora, a única filosofia que afirma que nós percebemos diretamente objetos e estados de coisas no mundo é o realismo direto, também conhecido como realismo ingênuo (*naïve*). Contudo, devemos fazer uma pequena distinção entre realismo direto e realismo ingênuo, pois, agora Searle reserva uma sutileza entre estes dois tipos de realismo. Em *Intencionalidade*, Searle (1983, p. 57) não estabelece esta distinção, ele tomava o realismo ingênuo e realismo direto como sinônimos, pois, em termos genéricos ninguém faria tal distinção. Atualmente, a título de distinção, o realismo direto é o que mais se

aproxima de sua teoria da intencionalidade perceptual. Para Searle, a teoria da percepção que endossa o realismo ingênuo é o disjuntivismo:

A visão da percepção que tenho defendido, isto é, que nós percebemos diretamente objetos e estados de coisas, muitas vezes é chamada de "*Realismo Direto*" e algumas vezes é chamada de "*Realismo Ingênuo*". É chamado de "realismo" porque diz que temos acesso perceptual ao mundo real, e "direto" porque diz que não precisamos primeiro perceber *alguma coisa a mais* por meio da qual percebemos o mundo real. Ela é muitas vezes contrastada com o Realismo Representativo, que diz que percebemos *representações do mundo real* em vez dos próprios objetos reais (SEARLE, 2015, p. 15).

Com relação a esse ponto, minha interpretação da teoria de Searle é que o seu realismo direto é uma consequência de sua explicação da intencionalidade perceptual e não o contrário. Searle pretende explicar como a intencionalidade perceptual funciona e, desde a publicação de *Intencionalidade*, essa intencionalidade perceptual pressupõe um tipo de causalidade intencional. Por conseguinte, esta causalidade intencional está relacionada a uma realidade objetiva, e esta realidade objetiva (ao menos para casos verídicos perceptuais) tem o poder de causar experiências subjetivas. Mesmo que seja questionável, a relação *apresentacional*<sup>B</sup> (*presentational*) dos objetos com o nosso campo visual subjetivo ocorrem, de acordo com Searle, de forma direta, e isto, como vimos acima, significa dizer que não “precisamos primeiro perceber alguma coisa a mais por meio da qual percebemos o mundo real” (SEARLE, 2015, p. 15).

Discutir o acesso direto à realidade é, em outras palavras, discutir a condição objetiva do mundo real, isto é, discutir a ideia de que, por um lado, o mundo é independente da minha vontade,<sup>4</sup> e, por outro, eu,

---

<sup>3</sup> Searle faz uma distinção entre representação (*representation*) e apresentação (*presentation*). Basicamente, a distinção gira em torno da ideia de que quando se trata de pensamentos sobre alguma coisa, temos casos representacionais, mas quando se trata de percepção de objetos reais no mundo, temos casos de apresentação. A percepção de uma maçã é uma apresentação da mesma e não uma representação. Cf. Searle (2015, p. 41).

<sup>4</sup> Podemos encontrar essa expressão no aforismo 6.373 do *Tractatus* de Wittgenstein, como sendo uma resposta ao solipsismo. Na página 109 do ensaio introdutório aos *Tractatus*, Luiz Henrique dos Santos traz a seguinte observação: “A vontade é independente dos fatos, mas não é independente da existência do mundo. A única condição da vida é que o mundo exista – não que exista este mundo, e não outro, mas que algo exista ao invés de nada. A existência do mundo tem o valor que tem a vida. A vida como vontade, deve estar em correlação essencial com a existência do

enquanto agente, sou causador de muitas mudanças neste mesmo mundo objetivo, logo, parece razoável estabelecer uma relação entre subjetividade e objetividade. Outra distinção<sup>5</sup> que Searle faz é entre o *epistemicamente objetivo* e *epistemicamente subjetivo*. Essa distinção está relacionada a questões relativas ao nosso conhecimento. Declarações epistemicamente objetivas lidam com questões de fato, enquanto que declarações epistemicamente subjetivas estão relacionadas a questões de opinião. Para elucidar esse ponto, Searle (2015, p. 16) usa como exemplo dois tipos de declarações envolvendo a figura do pintor Van Gogh. Se alguém afirma que Van Gogh morreu na França, esta afirmação é considerada epistemicamente objetiva, pois, a sua verdade ou falsidade pode ser entendida como uma questão de fato objetivo. Por outro lado, se alguém afirmar que Van Gogh foi melhor que Gauguin, esta afirmação será considerada epistemicamente subjetiva, pois, trata-se de uma questão de avaliação pessoal, neste caso, de uma preferência estética nas artes plásticas. De acordo com Searle, subjacente a essas distinções epistêmicas estão as distinções ontológicas.

Para entender minimamente a teoria da percepção de Searle, é preciso estar atento para uma distinção que ele faz entre *conteúdo* e *objeto*, que, em sua teoria, assume formas intencionais. Em termos explicativos, “duas experiências perceptuais podem ter conteúdos de mesmo tipo, mas uma tem objeto enquanto que a outra não” (SEARLE, 2015, p. 37). Searle chama de conteúdo intencional a experiência subjetiva que ocorre dentro de nossas cabeças, e de objeto intencional qualquer objeto pertencente ao mundo natural capaz de causar, em nós, experiências subjetivas perceptuais. A noção de causalidade perceptual é algo que foi influenciada pela teoria causal da percepção de Grice<sup>6</sup>, porém, em Searle, a causalidade

---

mundo, como valor. Tanto quando a vida como essência do pensamento, a vida como essência da vontade é inseparável do mundo. Com a morte, o mundo não se altera, mas termina.”

<sup>5</sup> Essa distinção também está presente na sua teoria da *Construção da Realidade Social*.

<sup>6</sup> Sobre isso, ver *The Causal Theory of Perception*. **Proceedings of the Aristotelian Society**. Supp. vol. xxxv. 1961. p. 121-53.

é intencional<sup>7</sup>. Com conceito de *auto-reflexividade causal*<sup>8</sup> (*Causal Self-reflexivity*), a “intencionalidade perceptual [...] tem como parte de suas condições de satisfação uma relação causal entre o estado intencional e o mundo externo” (SEARLE, 2015, p. 58). Desta forma, Searle só considera casos genuínos de percepção quando a experiência perceptual está ancorada em *condições de satisfação*, isto é, “o mundo será ou não será do modo como ele se parece para mim” (SEARLE, 2015, p. 57).

No próximo tópico, abordarei alguns aspectos problemáticos da teoria da percepção de Searle que estão diretamente ligados à relação entre conteúdo e objeto. Com essa relação, Searle acredita está desembaraçando algumas confusões conceituais cometidas pelos filósofos ao longo dos séculos. Todavia, veremos que no modo como Searle conduz o seu raciocínio, surgem algumas questões que dificultam compatibilizar seu realismo direto com a intencionalidade da percepção e a tese do elemento comum<sup>9</sup>.

## Representacionalismo e a tese do elemento comum

*Grosso modo*, as experiências perceptuais são divididas em três tipos: percepção verídica, ilusão e alucinação. No caso de percepção verídica, existe um objeto no mundo real responsável por causar minha percepção. Em termos searleanos, no caso verídico, o objeto do mundo real produz um conteúdo em minha experiência perceptual, a existência deste

---

<sup>7</sup> Searle (1983) chama a atenção de que, em algumas discussões da filosofia da mente, é muito comum haver uma distinção que aparentemente impossibilita uma relação entre a intencionalidade e a causalidade. A razão disto está em suas pressuposições originárias dos debates filosóficos em que a intencionalidade, muitas vezes, é conceituada como algo transcendente ou metafísico, que não faz parte do mundo natural. Em contrapartida, a causalidade é comumente entendida como uma relação natural de causa e efeito entre objetos e eventos físicos no mundo. Na concepção de Searle, devemos considerar a intencionalidade como parte do mundo natural e não como algo transcendente. Uma vez que a intencionalidade pode ser entendida como algo que faz parte do mundo físico, e as relações causais tradicionalmente são entendidas apenas como relações entre objetos físicos no mundo, Searle duplamente naturaliza a intencionalidade e intencionaliza a causalidade.

<sup>8</sup> Em *Intencionalidade*, esse conceito era usado por Searle pela denominação de *causalmente auto-referencial* (*Causally Self-referential*).

<sup>9</sup> Infelizmente, precisei suprimir muitos aspectos minuciosos da explicação da intencionalidade perceptual devido aos limites do texto, bem como ao seu propósito. Em outras palavras, os pormenores de como a intencionalidade perceptual funciona serão objeto de trabalhos futuros.

conteúdo é instanciada pela existência do objeto no mundo natural. Em outras palavras, na percepção verídica os objetos do mundo real são as causas das nossas percepções. No caso ilusório, o conteúdo intencional não corresponde de maneira verdadeira ao objeto no mundo real responsável por causar tal experiência. Mas para o caso ilusório, ainda assim existe um objeto no mundo real responsável pela percepção ilusória que pode ser ajustada (*direction to fit*) ou não a uma verdadeira e correspondente percepção do objeto. Por fim, o caso mais problemático entre os filósofos que é o caso alucinatório, neste caso, ocorre uma experiência visual dotada de conteúdo intencional embora sem um objeto no mundo real responsável por causar a experiência visual. Para as alucinações, a causa da fenomenologia da experiência visual não está em nenhum objeto do mundo natural, pois, esse objeto não existe. Para ilustrar esse ponto, poderíamos supor que alguém pudesse ter uma experiência visual de uma banana amarela (percepção verídica), ou que essa mesma banana pudesse ser vista na cor branca (ilusão), ou, até mesmo, poderia ser o caso de que não houvesse nenhuma banana (alucinação).

Em sua teoria, vimos que além de sustentar um realismo direto, Searle faz uma distinção entre objeto intencional e conteúdo intencional. Nos casos verídicos e alucinatórios de percepção, o conteúdo intencional pode ser de mesma natureza. Uma teoria disjuntivista da percepção estaria em acordo com o realismo direto e em desacordo com a ideia de que alucinações e percepções verídicas possuem a mesma natureza fenomenal. Para os disjuntivistas, existe uma diferença na natureza da experiência perceptual desses casos. O tipo de experiência que tenho quando percebo os objetos no mundo não é aquela que eu poderia ter se estivesse alucinando. Os disjuntivistas negam que experiências envolvidas em casos de percepção, ilusão e alucinação tenham a mesma natureza (SOTERIOU, 2014, 2016). Assim, uma das razões principais para que os disjuntivistas possam sustentar seu realismo ingênuo repousa sobre a negação da tese do elemento comum que Searle (2015, p. 166) chamou de *tese da comunalidade* (*commonality*). Para Searle, do ponto de vista da

fenomenologia, é perfeitamente concebível que o conteúdo de ambas as experiências seja da mesma natureza<sup>10</sup>.

A teoria disjuntivista da percepção é um bom contraponto ao realismo direto de Searle. Aqui, ela cumpre um duplo propósito: (1) Ela apresenta uma defesa do realismo direto que contrasta com a teoria de Searle na medida em que ela não endossa a tese do elemento comum. (2) A noção disjuntivista da percepção fortalece minha tese de que a explicação da intencionalidade perceptual de Searle é incompatível com sua defesa de um realismo direto, devido à tensão representacional em sua teoria quando esta pressupõe a tese do elemento comum. No início do texto, afirmei que, em linhas gerais, os principais aspectos da teoria da percepção de Searle são: (1) Uma defesa do realismo direto, (2) A tese de que a percepção é intencional e (3) A tese do elemento comum.

Em *Intencionalidade*, Searle (1983) não entrou no mérito de discutir o que ele agora chama de tese da *comunalidade*, tampouco, fez qualquer menção a uma teoria disjuntivista da percepção, que, por seu turno, já era algo bem familiar nas discussões filosóficas dos anos oitenta<sup>11</sup>. Por outro lado, em *Seeing Things As They Are*, Searle (2015) não só torna explícito que endossa a tese do elemento comum (*commonality*) como também dedica todo um capítulo do seu livro para discutir o disjuntivismo.

Por conseguinte, a questão que quero trazer agora é: como endossar a tese do elemento comum e defender um realismo direto? Disjuntivistas defendem um realismo direto/ingênuo e negam a tese do elemento comum. Representacionistas, via de regra, negam o realismo direto e defendem a tese do elemento comum. Nesse aspecto, Searle é um autor que se encontra dividido entre essas duas perspectivas no que concerne à explicação da percepção. Todo o seu esforço ao longo de *Seeing Things As They Are*, assim também como em *Intencionalidade*, é mostrar que é possível ser um realista direto e sustentar uma teoria da percepção

---

<sup>10</sup> Em sua interpretação da experiência de pensamento do cérebro numa cuba, Searle (2015, p. 157) afirma que “o experimento de pensamento faz sentido somente com a suposição de que a fenomenologia é exatamente a mesma nos dois casos”.

<sup>11</sup> As primeiras formulações de uma teoria disjuntivista da percepção remontam aos anos sessenta.

intencional comprometida com a tese do elemento comum. Tim Crane que possui uma visão intencionalista da percepção e endossa em sua explicação um tipo de representacionalismo (GERT, 2017), nega o disjuntivismo e defende a tese do elemento comum. De acordo com Crane (2001, p. 137), o realismo direto “falha em explicar o sentido em que a percepção e alucinação compartilham um caráter fenomenal.” De um ponto de vista representativo, Crane (2001, p. 136) considera que o realismo direto é uma explicação da percepção que basicamente afirma não existir uma relação em comum entre os estados mentais da alucinação e da percepção, a percepção envolve apenas uma genuína relação com objetos reais percebidos. Mas isso não parece se aplicar à teoria de Searle, pois, sua explicação pressupõe que é possível ser um realista direto e defender a tese do elemento comum a partir de uma explanação da intencionalidade perceptual.

Assim como Searle, Michael Martin (2009) endossa uma defesa de um tipo de realismo direto/ingênuo, porém, de um ponto de vista disjuntivo. Martin é um dos principais alvos de Searle em suas objeções ao que ele considera uma teoria disjuntivista da percepção. Um realista ingênuo, na concepção de Martin (2009, p. 273), tem que levar em consideração que nossa experiência sensorial do mundo deve ser, pelo menos em parte, não representacional. Aqui já temos algo que se distancia da abordagem de Searle, a saber, o caráter não representacional. Em sua teoria, Searle (2015, p. 41) usa o conceito de apresentação ao invés de representação, contudo, a apresentação para Searle (1983, p. 46) não passa de uma subespécie da representação. De acordo com Martin (2009, p. 272), uma abordagem disjuntiva da percepção deve ser considerada um ponto de partida para qualquer discussão sobre a natureza da experiência perceptual. Para uma teoria disjuntivista *alguns* de nossos episódios sensoriais são apresentações de uma realidade que existe independente da nossa experiência.

Uma teoria representativa também partilha da ideia de que a realidade existe independente de nossa experiência, esse é um pressuposto

comum para disjuntivistas e representacionalistas. Segundo Martin (2009, p. 273), o representacionista atribui conteúdo intencional ou representacional para explicar o caráter fenomenal das percepções e se refere a elas como “teorias intencionais da percepção”. Para ele, é comum em tais teorias a ideia de que nossa experiência sensorial é transparente, “que experimentalmente nos é apresentado um reino independente da mente e não um conjunto de qualidades ou entidades dependentes da mente cuja existência depende dessa consciência.” (MARTIN, 2009, p. 273). De acordo com Searle (2015, p. 59), o argumento da transparência foi usado contra a intencionalidade das experiências perceptuais visuais, mas tal argumento é, na verdade, apenas mais uma prova a favor de sua tese de que percebemos as coisas diretamente. Há na distinção de Searle entre os campos objetivos e subjetivos a ideia de que a percepção é transparente. Searle (2015, p. 107) afirma que não há nada no campo visual subjetivo que não seja ele mesmo objetivo. Na concepção de Martin (2009), o argumento da transparência é algo de comum em teorias representativas, mas Searle defende um tipo de realismo direto.

Martin considera que o realismo ingênuo é inconsistente com dois pressupostos que, segundo ele, são comuns em grande parte das discussões filosóficas da percepção. O primeiro ele chama de Naturalismo Experimental. O Naturalismo Experimental encerra a ideia de que “nossas experiências sensoriais, assim como outros eventos ou estados dentro do mundo natural, estão sujeitos à ordem causal [...] (ou seja, incluindo causas e condições neurofisiológicas) e psicológicas” (MARTIN, 2009, p. 274). A ideia por detrás disso é que a ciência pode manipular o cérebro de alguém e induzir, por exemplo, essa pessoa a uma experiência de alucinação. O segundo pressuposto é o que Martin chama de *a suposição do elemento comum* (*The Common Kind Assumption*) que, como já vimos, consiste na ideia de que percepção verídica e alucinação compartilham um elemento comum ou mesma fenomenologia. Segundo Martin (2009, p. 276),

uma maneira de ler a história da filosofia da percepção é vê-la em termos de um conflito entre o realismo ingênuo e o tipo de compromisso refletido nessas duas pressuposições.

De acordo com Searle, tanto o Naturalismo Experimental quanto a tese do elemento comum não implicam na negação do realismo direto. Na passagem abaixo, Martin expõe sua caracterização do Naturalismo Experimental juntamente com a suposição do elemento comum:

Assim, o Naturalismo Experimental impõe certas restrições ao que pode ser verdade em experiências alucinatórias. Tais experiências podem ter apenas objetos dependentes da experiência ou não ter nenhuma relação com objetos. Pela suposição do tipo comum, o que quer que seja verdade do tipo de experiência que se tem quando se está alucinando, o mesmo deve ser verdade para o tipo de experiência que se tem ao perceber. Portanto, ou a experiência de alguém quando percebe de forma verídica é de algum objeto dependente da mente, ou a experiência não é essencialmente uma relação com nenhum objeto (MARTIN, 2009, p. 275).

Searle (2015, p. 181) argumenta que é ambígua a noção de que qualquer coisa que seja verdade quando alguém está alucinando também deve ser verdade quando alguém está percebendo um objeto real. Ainda segundo Searle (2015), no que diz respeito ao caráter qualitativo ontologicamente subjetivo tanto do caso verídico quanto do caso alucinatório, é verdade que ambos compartilham de um mesmo conteúdo intencional. Por outro lado, isso também pode significar que

qualquer relação com os objetos que os dois casos tenham deve ser a mesma, e isso imediatamente produz um absurdo, porque o caso ruim (*bad case*) não é uma relação com um objeto (SEARLE, 2015, p. 181).

Searle está chamando a atenção para sua distinção de que, nos casos verídicos, a experiência visual possui um conteúdo intencional que foi causado por um objeto intencional, enquanto que uma alucinação possui um conteúdo intencional sem objeto intencional. Na visão de Searle

(2015), os filósofos confundem o conteúdo com o objeto e, muitas vezes, a alucinação em si é tomada por objeto da percepção. Logo, a suposição do elemento comum exige apenas que o caráter qualitativo seja o mesmo nos dois casos e que isso não é inconsistente com o fato de que um possui objeto e outro não.

Searle parte do pressuposto que é um *Bad Argument* achar que na alucinação existe um objeto como, por exemplo, um *sense-data*. Na caracterização de Martin (2009, p. 275) do *elemento comum*, se tomamos por verdade que o tipo de experiência que temos quando estamos alucinando é o mesmo tipo de experiência que temos quando percebemos objetos reais, então ou o objeto da experiência verídica de algum modo depende da mente, ou a experiência não é essencialmente uma relação com nenhum objeto. Segundo Searle, essa caracterização apresenta uma falha em distinguir conteúdo e objeto e isso significaria que o mesmo conteúdo implica o mesmo objeto. Para Searle (2015), o Naturalismo Experimental não tem como consequência que a tese do elemento comum entre o caso verídico e alucinatório implica que cada um tenha o mesmo objeto. A questão diz respeito apenas em distinguir casos que têm objetos dos que não têm objetos:

A conjunção do naturalismo experimental e a suposição de tipo comum não afirma que o mesmo conteúdo implica o mesmo objeto. Pelo contrário, a única razão pela qual nos concentramos nesses casos é que eles ilustram o ponto em que em duas experiências perceptivas você pode ter exatamente o mesmo conteúdo intencional com um objeto intencional em um caso e nenhum objeto intencional no outro caso (SEARLE, 2015, p. 182).

Segundo Martin (2009, p. 275), sustentar um realismo ingênuo com base nesses dois pressupostos é incorrer numa contradição. Searle compartilha de ambos os pressupostos. Para qualquer percepção que eu tenha de um objeto físico no mundo, devo supor que este mesmo objeto existe independente de minha percepção. Se eu partilho da tese do elemento comum, devo admitir a possibilidade de que a percepção que acredito ser de um objeto físico poderia ser uma alucinação, pois, em

princípio, suas ontologias são indistinguíveis. A pergunta agora seria: alucinações possuem uma existência que independe de nossa mente? Ora, se alucinações não são dependentes de objetos físicos, sendo suas causas de outra natureza, por que postular o elemento comum? Uma percepção de um objeto físico possui uma ontologia que depende do próprio objeto físico; no caso da alucinação, sua ontologia não depende de um objeto físico, em outras palavras, não é o objeto físico que sustenta sua ontologia. Para um disjuntivista que discorda da tese do elemento comum, é razoável supor que se trata de ontologias diferentes. E para Searle (2015), vimos que as ontologias não são diferentes, a diferença está apenas no fato de que a percepção verídica possui um objeto que causa a experiência visual, enquanto que na alucinação as causas são desconhecidas e não possuem um objeto como sustentáculo da experiência visual.

Na tentativa de situar o realismo direto de Searle, proponho aqui fazermos uma distinção entre em dois sentidos para “percepção direta”, um sentido forte e um sentido fraco. Por sentido fraco, quero expressar a ideia de que nossa relação perceptual com o mundo ocorre de forma direta e imediata. Isso significa que não percebemos dados dos sentidos, nem representações ou ideias. Por outro lado, a linguagem, ao menos para nós humanos, é um meio segundo o qual podemos discernir e categorizar conceitualmente nossas percepções. Searle (1983, p. 54) afirmou que “muitas de nossas experiências visuais nem sequer são possíveis sem o domínio de certas capacidades de *background* e proeminente entre elas são as capacidades linguísticas”. Ora, não seria a linguagem ela mesma um meio segundo o qual eu acesso os objetos da percepção? Em *Intencionalidade*, Searle (1983, p. 54) usa o exemplo do signo “TOOT”, que, a depender do *background* de capacidades, podemos interpretar perceptualmente este signo de vários modos. Como relacionar estas variadas respostas com a ideia de que percebemos diretamente? Mas aqui estamos falando do sentido fraco de percepção direta. O que dizer do sentido forte de percepção direta? Se quisermos ser rigorosos, percepção direta no sentido forte não é possível, e se ao menos pudermos conceber

alguma possibilidade, ainda assim será problemático. O mais próximo que poderíamos chegar disso seria apelar para uma definição de Searle chamada percepção básica, a saber, “qualquer percepção de um objeto ou característica que você possa ter sem perceber algo mais através da qual você a percebe” (SEARLE, 2015, p. 112).

Em Searle (2015, p. 112), a definição de percepção básica está ligada apenas ao que ele chama de características de nível mais baixo que aqui podemos resumir por cores e formas. Penso que cores e formas encerram capacidades conceituais de nossa percepção que, por sua vez, estão ancoradas em capacidades linguísticas. Entretanto, desde *Intencionalidade*, Searle (1983) parece oscilar entre seu realismo direto e uma tácita teoria representacional da percepção. Minha interpretação é que isso pode ser traduzido como uma inconsistência entre sua tese de que a percepção é intencional e sua tese de que percebemos as coisas diretamente, pressupondo nesse caso uma defesa do realismo direto.

Apesar da atual teoria da percepção de Searle possuir um arcabouço conceitual mais amplo, ele ainda continua incorrendo em dificuldades de compatibilizar sua tese de que a percepção é intencional e que nosso acesso à realidade é direto. A razão disso repousa, como vimos, sobre a suposição de um elemento comum entre a fenomenologia da percepção do caso verídico e alucinatório. Nesse sentido, podemos suspeitar que o realismo direto de Searle pudesse ser interpretado como uma forma de representacionalismo. Na explicação de Searle (2015, p. 106) sobre os campos visuais da percepção, ele afirmou que no campo visual objetivo tudo é visto ou pode ser visto, enquanto que o campo visual subjetivo, nada é visto nem pode ser visto. A pergunta que surge é: o que Searle pretende com isso? Dado que no campo visual subjetivo nada é visto, isso não significa que “as entidades no campo visual subjetivo são invisíveis, mas sim que, sua existência é o ver dos objetos no campo visual objetivo” (SEARLE, 2015, p. 107). Searle afirma que quando vemos alguma coisa, não vemos o “ver da coisa” (campo visual subjetivo), o que vemos são os conteúdos intencionais que podem ou não ter objetos intencionais, como

é o caso da alucinação que possui um conteúdo intencional, mas carece de um objeto intencional. Para os casos alucinatorios, Searle vai dizer que a pessoa que sofre de uma alucinação não vê<sup>12</sup> nada, mas ainda assim possui um conteúdo intencional ancorado num tipo de experiência visual. Penso que esse ponto da explicação de Searle pode trazer alguns problemas.

De todas as considerações que tivemos até agora sobre a teoria da percepção de Searle tanto em *Intencionalidade* quanto em *Seeing Things As They Are*, é possível, a partir do seu comprometimento com a tese do elemento comum, perceber uma tensão representativa em sua explicação. Minha interpretação é que se Searle quiser ser coerente, ele deve fazer uma escolha: ou sacrificar seu realismo direto e admitir uma teoria intencionalista representativa da percepção, ou sacrificar sua explicação da intencionalidade perceptual e adotar um tipo de realismo ingênuo que nega a tese do elemento comum. Entretanto, a teoria da intencionalidade da percepção de Searle pressupõe o realismo direto e não o contrário. Fazendo uma conjectura, acredito que se Searle pudesse abrir mão de uma dessas duas alternativas, ele sacrificaria seu realismo direto, pois a intencionalidade da percepção ocupa um lugar de maior privilégio em sua filosofia, o seu comprometimento com a tese do elemento comum também é um indicativo disso. Além do mais, uma teoria representacional é muito mais comum e compatível com o modelo científico vigente do que o realismo direto. Em *Seeing Things As They Are*, há uma passagem de sua discussão com os disjuntivistas no qual Searle preferiu se aliar a Descartes (seu arqui-inimigo filosófico) do que abrir mão da tese do elemento comum. Essa passagem também é mais um indicativo de seu tácito representacionalismo:

---

<sup>12</sup> O ver (*seeing*) para Searle só ocorre em casos verídicos e ilusórios de percepção pelo fato deles estarem causalmente conectados com algum estado de coisas no mundo. Nesse sentido, nos casos de alucinação não existe o ver (*seeing*) da percepção, pois como já foi enfatizado, em casos alucinatorios não há um objeto da percepção, a alucinação não pode ser tomada como objeto.

Eu, juntamente com Descartes e praticamente todos os demais, por estipulação decido considerar exatamente iguais casos de experiência verídica e casos de alucinações correspondentes (SEARLE, 2015, p. 170).

Descartes era um representacionista que defendia que o tipo de acesso que temos à realidade era descrito em termos de uma teoria da representação e não de um realismo direto. Em sua defesa, Searle poderia responder que ele compartilha da tese cartesiana do elemento comum, mas que rejeita a forma como acessamos os objetos da percepção. A experiência perceptiva para Searle é sempre uma apresentação mental interior que é causada por estados de coisas do mundo. Além do mais, o conteúdo de nossas experiências perceptivas é intrínseco a essas experiências e não aos objetos exteriores. Percebam que nesse modelo explicativo há um jogo dualista de interior e exterior. Ao longo de nosso texto, vimos várias distinções que apontam para esse dualismo: objetividade e subjetividade; ontologicamente objetivo e ontologicamente subjetivo; epistemologicamente objetivo e epistemologicamente subjetivo; conteúdo intencional e objeto intencional etc. Nesse jogo dualista, Searle afirma haver uma distinção entre o conteúdo das experiências perceptivas e as propriedades físicas dos objetos externos.

O meu ponto aqui não é discutir se uma teoria da percepção com base no realismo direto explica melhor o problema da percepção do que uma teoria representativa. Ambos os lados enfrentam problemas. Minha questão é que a teoria de Searle, que tem como base a intencionalidade e a defesa do elemento comum, parece incorrer em uma forma de representacionismo.

No próprio modelo explicativo de Searle, quando ele faz uma distinção entre conteúdo intencional e objeto intencional, poderíamos interpretar essa distinção como uma forma de replicar os objetos do mundo externo em conteúdos intencionais do mundo interno. Afirmar que existe um objeto externo que causa minha experiência interna (conteúdo intencional) e que esse objeto não deve ser confundido com o conteúdo não é muito diferente da ideia representacionista em que não devemos

confundir os objetos internos não físicos que percebemos direta e imediatamente, com os objetos físicos do mundo externo que percebemos indiretamente. E de fato não podemos confundir, até porque para Searle o caráter fenomenal de uma percepção verídica pode ter a mesma natureza de uma alucinação. Para ilustrar minha interpretação, vou recorrer a algumas definições básicas para realismo direto e representacionalismo e incluir uma definição disjuntivista juntamente com o que seria uma definição searleana:

- (1) *Realismo indireto ou representacional*: percepção indireta de objetos físicos e estados de coisas no mundo através da percepção direta e imediata de seus estados mentais representacionais.
- (2) *Realismo direto*: percepção direta e imediata de objetos físicos e estados de coisas sem perceber nada de intermediário no processo perceptual.
- (3) *Realismo ingênuo disjuntivista*: percepção direta e imediata de objetos físicos e estados de coisas. O objeto da percepção é o próprio estado de coisas cujo caráter fenomenal é distinto do caráter fenomenal de uma alucinação.
- (4) *Realismo de Searle*: percepção direta e imediata de objetos físicos e estados de coisas, mas com uma diferença entre conteúdo intencional e o objeto intencional, além de uma defesa da tese do elemento comum.

Se formos rigorosos, o realismo de Searle não está em (1) nem em (2), tampouco em (3). Sua defesa de um realismo direto seria algo à parte de qualquer definição de realismo direto, dado sua minuciosa descrição do fenômeno da intencionalidade perceptual. Por outro lado, conjecturei que se Searle tivesse inevitavelmente que escolher entre sua explicação intencional da percepção e sua defesa do realismo direto, seguramente ele ficaria com a intencionalidade da percepção, embora saibamos que, na prática, isso seria inadmissível. Minha pergunta agora é: a partir dessas caracterizações feitas acima, em qual das três primeiras a teoria de Searle mais se enquadra? Na introdução desse trabalho, lancei mão da pergunta que indiretamente norteou todo o texto, a saber: é consistente a explicação intencional da percepção em Searle com a sua defesa de um realismo direto? Em seguida, afirmei que a pergunta pode ser formulada tanto em

*Intencionalidade* quanto em *Seeing Things As They Are*. Agora, novamente, estou tentando argumentar que apesar da teoria de Searle possuir um forte apelo intuitivo e explicativo acerca da tese de que a percepção é intencional e direta, ela sustenta um comprometimento representacionalista.

Voltemos à distinção de Searle entre conteúdo intencional e objeto intencional. Podemos interpretar essa distinção como uma forma de replicar os objetos do mundo externo em conteúdos intencionais do mundo interno. A causação intencional apresenta no meu campo visual subjetivo um conteúdo intencional que tem como origem causal um objeto físico do mundo real. Eu não veria um objeto diretamente, o que veria seria um conteúdo intencional que, a depender do caso, pode ter uma *forma aspectual* (*aspectual shape*) bem específica, segundo o qual outra pessoa poderia perceber algo completamente diferente dado seu *background* de capacidades. Se isso é possível, então a própria teoria de Searle também seria um tipo de *Bad Argument*. Se eu percebo uma *forma aspectual* e atribuo esta forma a um objeto externo, por que não vejo o próprio objeto externo? A obra de Searle se chama *Seeing Things As They Are*; mas será que vemos de fato as coisas como elas são? O “*ver*” (*seeing*) para Searle é uma experiência subjetiva que ocorre dentro de minha caixa craniana como no exemplo de sua interpretação do cérebro numa cuba. Toda e qualquer experiência visual é sempre de um conteúdo intencional, a diferença é que, na explicação de Searle, alguns conteúdos possuem uma causa determinada enquanto que outros não. Em uma passagem sobre o que postula o realismo direto, ele afirma que “a experiência visual é o conteúdo e não o objeto da percepção”. (SEARLE, 2015, p. 173). A pergunta aqui agora seria: como posso ter certeza de que um objeto causou em mim uma experiência visual, se o que eu tenho acesso é o conteúdo intencional com suas características intrinsecamente qualitativas? Se eu vejo os objetos diretamente, porque distinguir entre conteúdo e objeto?

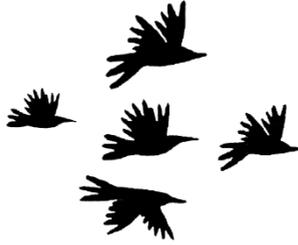
Ao sustentar a tese do elemento comum, penso que Searle estaria sustentando um tipo de representacionismo. Por mais elegante que seja

sua teoria, por mais amplo que seja seu arcabouço conceitual para explicar que a percepção é direta, não significa que, de fato, percebemos as coisas diretamente. As dicotomias conceituais explicativas, juntamente com a tese do elemento comum, parecem mostrar que não percebemos diretamente objetos e estados de coisas no mundo. Nesse caso, acreditar que percebemos os objetos e estados de coisas diretamente, não significa que de fato percebemos objetos e estados de coisas diretamente. Em *Intencionalidade*, todas as vezes que o ceticismo ameaçava o modelo explicativo de Searle (1983, p. 58 e p. 76), ele se esquivava da discussão simplesmente afirmando que não estava interessado em discutir os problemas clássicos da filosofia. Por outro lado, em *Seeing Things As They Are*, Searle (2015, p. 220) afirmou que o ceticismo – que ameaçava pensadores como Descartes, Locke, Berkeley, Hume e Kant – não ameaçava sua teoria. Se podemos fazer uma leitura representacionista de Searle, ao passo que percebemos resquícios cartesianos em sua explicação, um representacionista que olhasse com atenção para suas teses perceberia que Searle estaria sendo o pior dos cartesianos, pois o seu cartesianismo é disfarçado. Para despistar esse cartesianismo, Searle simplesmente postulou que acessamos o mundo diretamente e que nada está intermediando nossa relação perceptual com o mundo. Por outro lado, em termos filosóficos, existe uma obsessão de Searle em refutar qualquer coisa que assuma uma forma dualista cartesiana, e isso desde as suas considerações em filosofia da mente, quando ele afirmou que as explicações filosóficas da mente repousam num *dualismo conceitual*. Quando se trata do trabalho de Searle, esta tarefa parece ser infundável. Em sua perene batalha contra a hidra cartesiana, para cada cabeça cortada, outras voltam a crescer.

## Referências

CRANE, Tim. **Elements of mind**: An introduction to the philosophy of mind. USA. Oxford University Press, 2001.

- CRANE, Tim. **The problem of perception.** 2015. Disponível em: <https://plato.stanford.edu/entries/perception-problem/>. Acesso em: 08/07/2018.
- GERT, Joshua. **Primitive colors:** A case study in neo-pragmatist metaphysics and philosophy of perception. Oxford University Press, 2017.
- GRICE, H. P. The causal theory of perception. **Proceedings of the Aristotelian Society**, Supp. vol. xxxv, 1961, p. 121-53.
- MARTIN, Michael G. F. The limits of self-awareness. *In*: BYRNE, Alex; LOGUE, Heather (Ed). **Disjunctivism:** Contemporary Readings. Massachusetts. The MIT Press, 2009, p. 271-317.
- MAUND, Barry. **Perception.** Chesham: Acumen, 2003.
- SEARLE, John. **Intentionality:** An essay in the philosophy of mind. Cambridge University Press, 1983.
- SEARLE, John. **The Rediscovery of the mind.** Massachusetts: The MIT Press, 1992.
- SEARLE, John. **Mind:** A brief introduction. Oxford University Press, 2004.
- SEARLE, John. Perceptual Intentionality. **Organon F**, 19, 2012, p. 9-22.
- SEARLE, John. **Seeing things as they are:** A theory of perception. Oxford University Press, 2015.
- SOTERIOU, Matthew. **The disjunctive theory of perception.** 2014. Disponível em: <https://plato.stanford.edu/entries/perception-disjunctive/>. Acesso em: 17/11/18.
- SOTERIOU, Matthew. **Disjunctivism.** New York: Routledge, 2016.
- WITTGENSTEIN, L. **Tractatus logico-philosophicus.** Trad. de Luiz Henrique dos Santos. São Paulo: Edusp, 2008.



## Relações causais e racionais entre mente e mundo: a resposta de Donald Davidson a John McDowell

*Marco Aurélio Sousa Alves*<sup>1</sup>

### Introdução

A interação entre nossos pensamentos e o mundo deve dar conta de pelo menos duas questões distintas, mas inseparáveis. De um lado, a questão *epistemológica*: como nossas crenças, nossa visão de mundo, podem ser justificadas, levando-nos a um conhecimento do mundo. De outro lado, a questão *transcendental*: como nossos pensamentos podem se dirigir ao mundo, falar dele, tomar partido frente a como as coisas são. O primeiro problema enfrenta o desafio cético, que se instala na fissura aberta entre o domínio interno do pensamento e o mundo externo. O segundo busca acomodar a intencionalidade, ou a direção ao mundo de nossas atividades mentais. Uma compreensão satisfatória da relação entre pensamento e mundo precisa oferecer respostas a ambas as questões.

Quando publicou *Mente e Mundo*, em 1994, John McDowell instaurou um debate com Donald Davidson que irá se estender até a morte deste último, em 2003. O debate orbita em torno da crítica que McDowell

---

<sup>1</sup> Professor Adjunto - DFIME/UFSJ; Professor Permanente - PPGFIL/UFSJ e POSDEFIL/UFOP.

endereça a Davidson, segundo a qual este teria ignorado a questão propriamente transcendental e se preocupado exclusivamente com o aspecto epistemológico da relação entre mente e mundo. Tal “cegueira transcendental” teria levado Davidson a defender uma visão incapaz de explicar o fato de nossos pensamentos dizerem respeito ao mundo. Parte dessa deficiência, segundo McDowell, estaria relacionada à estratégia davidsoniana de admitir apenas transações *causais* entre mente e mundo.

No que se segue, não irei esmiuçar os vários elementos da crítica de McDowell a Davidson.<sup>2</sup> Irei me concentrar exclusivamente na acusação de que Davidson, ao aceitar apenas transações causais entre mente e mundo, teria tornado o mundo incapaz de oferecer razões para pensarmos como pensamos. Primeiramente, reconstruirei alguns dos elementos que conformam a perspectiva de Davidson, que remontam, fundamentalmente, a Quine e Sellars. Em seguida, veremos como Davidson irá se apropriar desses elementos e formular a perspectiva que será alvo da crítica de McDowell. Tal crítica, conforme pretendo mostrar, não faz jus à forma como Davidson compreende a relação entre causas e razões. Após explicitar tal relação, argumento que a crítica de McDowell, nesse ponto específico, mostra-se injustificada e apressada.

## A herança de Quine

Quine (1963) criticou a ideia de que poderíamos detectar isoladamente, termo a termo ou enunciado a enunciado, o significado empírico de expressões linguísticas. Trata-se de sua famosa crítica ao *reducionismo*, um dos dogmas do empirismo. A significação empírica, não podendo ser detectada nas partes, deveria ser buscada no conjunto, na forma como nossa visão de mundo se relaciona com a realidade.

O dogma do reducionismo deve ser compreendido na companhia de outro dogma: a *distinção entre analítico e sintético*. Ao olharmos para cada enunciado isoladamente, não podemos dizer o quanto sua verdade ou

---

<sup>2</sup> Para um mapeamento detalhado desse debate, ver Alves (2008).

falsidade depende do mundo ou da própria linguagem. Os dois elementos determinam, ao mesmo tempo, a verdade de todos os enunciados, só não podemos identificar essa contribuição caso a caso. Nem podemos distinguir enunciados sintéticos de analíticos, ou seja, aqueles que dependem em parte do mundo e aqueles que em nada dependem dele para sua verdade ou falsidade.

Frente à rejeição dos dois dogmas, Quine (1963, p. 41) propõe que vejamos nossos enunciados enfrentando o “tribunal da experiência sensível não individualmente, mas como um corpo coeso”.<sup>3</sup> Quanto à questão da intencionalidade, a como nossa linguagem se liga ao mundo e, portanto, pode falar dele, Quine (1963, p. 42) mantém a exigência empirista de um contato com o mundo através da experiência, mas ressalta, tendo em vista o abandono dos dogmas, que “a unidade da significação empírica é a ciência como um todo”. Com isso, ele mantém o jogo entre elementos linguísticos e extralinguísticos na constituição da significação, só não admitindo a verificação excessivamente fina da contribuição desses elementos nas partes da linguagem.

Quine não abandona, portanto, a ideia de conteúdo empírico. Ele (1963, p. 43) ressalta apenas que “é enganoso falarmos em conteúdo empírico de um enunciado individual”. Vista como um bloco, a linguagem não só possui como precisa possuir tal conteúdo: “tomada como um todo, a ciência tem essa dupla dependência da linguagem e da experiência” (QUINE, 1963, p. 42). O termo ‘ciência’ pode ser traduzido aqui, sem violência ao sentido original, por ‘visão de mundo’. Já a palavra ‘linguagem’ se refere ao que é interno, ao nosso *esquema conceitual*, àquilo que independente da experiência. Nossos conceitos são cunhados conforme considerações pragmáticas: COELHO, MESA ou NEUTRINO, ou seja, conceitos ordinários ou científicos, são pragmaticamente moldados para lidarmos com o fluxo da experiência. Esquemas conceituais possuem algo de arbitrário e acidental, ou melhor, refletem algo dos nossos

---

<sup>3</sup> Todas as traduções de Quine são de minha autoria.

interesses e daquilo que fazemos, podendo assumir formas radicalmente diferentes. Mas, ao mesmo tempo, eles mantêm uma relação com o mundo. A experiência atua de forma a tornar nossa visão de mundo uma construção, holisticamente organizada, de elementos internos e externos ao esquema conceitual.

Uma pergunta crucial parece acompanhar todo o pensamento de Quine (1974, p. 1): “dadas apenas as evidências dos sentidos, como chegamos à nossa teoria do mundo?”. Investigar isso, segundo sua visão naturalizada, é a finalidade mesma da epistemologia. As “evidências dos sentidos” são compreendidas como uma forma de acesso indireto à realidade lá fora. Quine (1997, p. 1) afirma, por exemplo, que “objetos físicos em geral, não importa o quão distantes, tornam-se conhecidos para nós apenas através dos efeitos que ajudam a induzir em nossas superfícies sensórias”. Essa entrada sensória é bastante pobre, ao contrário de nossa visão de mundo, que é riquíssima. Como explicar esse descompasso, ou “a relação entre o *input* pífio e o *output* torrencial?” (QUINE, 1969, p. 83).

A resposta está na contribuição humana na construção de sua visão de mundo. A experiência, vista como impactos sensórios, desencadeia um processo no qual o ser humano, livremente, movido por considerações pragmáticas, faz o restante do trabalho. O resultado é nossa visão de mundo. O mundo fornece apenas algumas *dicas*, “subtraindo as dicas de sua visão de mundo, temos a contribuição humana como a diferença. Essa diferença marca a extensão da soberania conceitual do homem” (QUINE, 1997, p. 5).

A experiência desempenha, portanto, um papel fundamental: ela limita a extensão da soberania conceitual. Sem ela, estaríamos totalmente divorciados do mundo. Sem a experiência, não teríamos propriamente uma visão *de mundo*, não poderíamos falar dele, ou tê-lo em mente. O empirismo sobrevive em Quine, livre dos dogmas, porque essa questão transcendental, acerca da possibilidade de termos o mundo em vista, ficaria misteriosa sem o papel controlador da experiência.

A leitura que John McDowell faz de Quine enfatiza adequadamente o papel transcendental desempenhado pela experiência em seu pensamento:

De acordo com essa concepção, o fato de algo (uma crença, por exemplo, ou, para ficarmos mais próximos ao espírito do texto quineano, a totalidade de uma visão de mundo) dizer respeito ao mundo, no sentido de constituir uma tomada de posição sobre o modo como as coisas são, depende de esse algo ser vulnerável ao mundo no que diz respeito a um veredicto sobre sua aceitabilidade. Esse veredicto, por sua vez, só pode ser dado através da experiência (McDOWELL, 2005, p. 169).

Quine usa a imagem do “tribunal da experiência” exatamente para mostrar como a experiência, conforme a velha e boa tradição empirista, julga as nossas crenças, constringendo ou conformando nossa visão de mundo. Diante dos veredictos da experiência, são realizados ajustes holisticamente arranjados na estrutura de nossa visão de mundo. Apesar da relação um tanto indireta entre experiência e crenças específicas, podemos dizer que há um confronto entre nossa teoria e o mundo, cabendo à experiência o papel de nos indicar se devemos ou não alterar alguma coisa em nossa visão de mundo.

Davidson foi enormemente influenciado pelo pensamento de seu antigo mestre, Quine. Apesar do tanto que seus pensamentos têm em comum, iremos nos deter aqui em alguns pontos fundamentais nos quais eles divergem. Ao contrapor esquema conceitual e mundo, Quine precisou introduzir intermediários epistêmicos capazes de transportar as “entregas dos sentidos” (*deliverances of the senses*) e dotar a linguagem de conteúdo empírico. Davidson acredita que essa imagem apresentada por Quine é vulnerável ao ceticismo. Sua crítica se dirige especificamente ao apelo a “intermediários epistêmicos”. Segundo Davidson, tal noção implica uma separação entre rede de conceitos e mundo externo: “o problema familiar

é, evidentemente, que a desconexão cria um fosso que nenhum raciocínio ou construção pode transpor” (DAVIDSON, 1988, p. 43).<sup>4</sup>

O problema da contraposição entre conceitos e dados brutos vindos do mundo reside, para Davidson, na falta de garantias epistemológicas. Se almejamos um conhecimento com pretensões objetivas, não podemos nos fiar em evidências subjetivas. As sensações, entregas dos sentidos, irritações nas terminações nervosas, dados sensoriais, ou qualquer outro análogo, são ocorrências privadas. Tais ocorrências não podem servir de base para a justificação de um conhecimento objetivo do mundo. A dúvida cética acerca da impossibilidade de alçarmos uma posição na qual estaríamos autorizados a falar do mundo tal como ele é se mostra cogente para quem aceita os termos dessa contraposição.

Mesmo se ignorarmos o caráter privado da evidência, teríamos ainda dificuldade para explicar a justificação de nossa visão de mundo a partir de intermediários epistêmicos. A existência de um intermediário já é suficiente para ameaçar a possibilidade de o mundo se apresentar diretamente à mente. Podemos, junto com os céticos, perguntar o que garante a confiabilidade da informação transmitida. Nossa visão de mundo se basearia, dessa forma, em mensageiros que nos dão dicas acerca de como são as coisas lá fora, e só podemos avaliar as informações de um mensageiro a partir das informações de outros mensageiros. Estaríamos condenados a confiar em intermediários sem termos como verificar diretamente a veracidade do que dizem. Em tal situação, podemos ser sistematicamente enganados por gênios malignos sem nunca notarmos.

O dualismo que será alvo da crítica de Davidson opõe rede conceitual e realidade, e depois tenta ligar um ao outro através dos dados brutos da experiência. Mas a experiência sensorial impõe intermediários que não podem mais garantir nada. Ficamos, assim, com uma certeza subjetiva que nunca consegue se tornar saber objetivo: situação análoga à de Descartes, mas sem um deus veraz que nos proteja do gênio maligno. Davidson acredita que continuamos a pensar, com esse dualismo, segundo o modelo

---

<sup>4</sup> Todas as traduções de Davidson são de minha autoria.

cartesiano, e o resultado disso é uma visão de mundo que não consegue explicar sua objetividade.

Mesmo em sua expressão mais sofisticada, que Davidson detecta em Quine, o dualismo entre esquema conceitual e realidade exige sempre alguma forma de intermediário epistêmico. Necessita dele para ligar um polo ao outro. O esquema precisa de conteúdo empírico, e isso só pode ser fornecido pela experiência. Mas a experiência é subjetiva, e o mundo continua a rodar fora de nossos domínios. Os intermediários não cumprem, portanto, a promessa de nos ligar à realidade. Davidson entendeu por *empirismo* a tese de que as entregas dos sentidos são epistemologicamente significativas. O empirismo postula evidências privadas para as crenças, incorrendo, portanto, num tipo de subjetivismo tipicamente moderno. Mesmo o empirismo sem dogmas de Quine incorreria nesse equívoco.

O dualismo detectado em Quine é apresentado por Davidson (1974, p. 189) como sendo um novo dogma do empirismo; “o terceiro e, talvez, último dogma”. O dogma consiste em atribuir às entregas dos sentidos o duplo papel de controlar nossa soberania conceitual e tornar nossa visão de mundo justificada. Segundo Davidson, tal dogma é problemático pois mantém o caráter subjetivo das evidências sensoriais e não consegue, assim, evitar o ceticismo tipicamente moderno acarretado pela introdução de intermediários epistêmicos. Essas críticas levam Davidson a buscar uma alternativa que elimine completamente o projeto empirista de fundamentar o conhecimento do mundo e o conteúdo empírico a partir da experiência sensível.

### **A herança de Sellars**

As críticas que Davidson faz ao empirismo quineano, bem como a alternativa que ele irá defender, são enormemente influenciadas pela famosa crítica de Wilfrid Sellars (1956) ao mito do dado. Podemos formular a crítica ao mito do dado da seguinte forma: a experiência,

entendida como um fenômeno natural, fora do “espaço das razões”, não pode fornecer, dar ou inserir nada no interior do jogo de justificações. Vista como puramente sensível, a experiência não pode estabelecer qualquer relação racional com as crenças.

Sellars (1956, p. 76) verifica que quando caracterizamos certos estados ou episódios como sendo um conhecer (*knowing*), estamos “situando essas coisas no espaço lógico das razões, de justificar e ser capaz de justificar o que se diz”<sup>5</sup>. É traçada, assim, uma linha distinguindo dois espaços. As caracterizações que pertencem ao espaço das razões são irreduzíveis ao espaço da natureza. Tentar descrever acontecimentos do primeiro espaço em termos do segundo nos levaria a incorrer numa espécie de falácia naturalista em epistemologia. O que se procura vetar é a possibilidade de reduzirmos, sem sobras, fatos epistêmicos a fatos não-epistêmicos ou naturais.

Quando fala em “dado”, Sellars tem em mente a noção de dado sensível (*sense datum*). Tal noção é vista como um híbrido oriundo de duas linhas distintas de pensamento. De um lado, episódios sensórios são entendidos como não envolvendo conceitos. O exemplo clássico é a sensação das cores. Do outro lado, temos conhecimentos não-inferenciais de que tal-e-qual é o caso. Cruza-se, assim, a fronteira entre os dois espaços lógicos. Uma coisa é a mera sensação bruta, outra a apreensão intensional de um elemento do mundo. A segunda é conceitualmente estruturada e pode habitar o espaço das razões. A primeira é uma relação descrita em termos naturais e não pode justificar ou ser razão para nada. Grande parte do esforço de Sellars consistiu exatamente em formular uma imagem aceitável de como o sensível e o conceitual se combinam de forma a permitir a intencionalidade da experiência perceptiva.

A crítica de Sellars acusa tal noção de operar a seguinte manobra fraudulenta: uma sensação bruta não pode *justificar* coisa alguma. Temos aí uma mistura de privado e público, conceitual e não-conceitual, fatos epistêmicos e fatos naturais. A passagem de um lado para o outro dessa

---

<sup>5</sup> Todas as traduções de Sellars são de minha autoria.

fronteira não pode ser pensada como uma relação racional. Aí reside o mito. Sellars insistiu que qualquer projeto que tente tornar inteligível uma noção anfíbia de experiência, que se situe entre o mundo não-conceitual e nossa linguagem conceitualmente estruturada, está fadado ao fracasso<sup>6</sup>.

Tendo em vista a crítica ao mito do dado, salta aos olhos o caráter problemático da imagética jurídica de Quine. O tribunal da experiência não poderia julgar nada sem introduzir, sub-repticiamente, dados exógenos em nossa rede de crenças. Emitir veredictos é, evidentemente, uma relação racional que a experiência supostamente travaria com o nosso sistema de crenças. Se a crítica ao dado implica a impossibilidade de uma transação racional entre experiência e crença, então a imagética quineana é insustentável. Ela pretendia explicar como nossa visão de mundo é racionalmente constrangida pela experiência, mas tenta fazer isso com uma manobra fraudulenta.

McDowell detecta em Quine uma oscilação constante entre dois significados distintos do termo ‘experiência’ (*experience*). Concebida como mera causação bruta em nosso aparato sensível, a experiência não pode constituir um movimento no interior do espaço das razões. Quando Quine (1969, p. 75) afirma, por exemplo, que “a estimulação dos receptores sensíveis é toda a evidência que alguém possui, no fim das contas, para chegar à sua imagem do mundo”, certamente ele tem essa noção bruta de experiência em mente. E nessa passagem vemos ele escorregar, sub-repticiamente, do exterior para o interior do espaço das razões. Nesse sentido, a crítica de Sellars se mostra procedente. McDowell (2005, p. 135) detecta também um segundo sentido de ‘experiência’, vista não como “a estimulação de nossos receptores sensíveis”, mas sim como “o modo como as coisas nos aparecem”. Entendida da segunda forma, seria aceitável dizer que a experiência oferece o material necessário para chegarmos à nossa visão de mundo. O problema, como percebeu Davidson e apontou

---

<sup>6</sup> Ver Brandom (2002).

adequadamente McDowell, é que Quine oscila de um sentido ao outro e acaba sujeito à crítica de Sellars ao mito do dado.

Davidson percebeu muito bem o movimento inadequado operado pela noção de experiência em Quine. Para evitar tal problema, ele viu que era preciso distinguir claramente relações causais de relações racionais. A proposta que Davidson irá chamar inicialmente de coerentista pretende exatamente eliminar essa confusão, tirando a experiência definitivamente da ordem das justificações. Em sua proposta coerentista, não é permitido à experiência emitir veredictos, justificar ou julgar qualquer coisa. A experiência, em suma, não estabelece uma relação racional com as crenças. Entendida como um fato propriamente natural, a experiência interage apenas causalmente com nosso sistema de crenças. A justificação é um processo que se dá no interior do sistema de crenças, consistindo num tipo de relação racional ou epistêmica que crenças estabelecem entre si. Não há algo vindo de fora do sistema de crenças, um dado bruto ou experiência sensível, que seja capaz de interferir nesse jogo de justificações. Daí a famosa máxima coerentista de Davidson (1983, p. 141): “nada pode contar como uma razão para sustentar uma crença exceto uma outra crença”.

O fato de apenas crenças (ou, em sentido mais amplo, estados mentais intencionais) poderem se relacionar racionalmente entre si deixa a relação entre pensamento e mundo restrita a uma relação de tipo causal. Uma vez que a experiência é compreendida como mera impressão sensível, ela não pode justificar nada. O círculo epistêmico é fechado em seu próprio domínio. Temos apenas uma abertura causal, a partir da qual o mundo causa sistematicamente as nossas crenças. O problema evidente do coerentismo ingênuo era explicar como nosso sistema de crenças, ainda que coerente, poderia se relacionar com o mundo objetivo. Davidson buscou contornar essa dificuldade apresentando uma forma alternativa de relacionar nossas crenças (ou atividades mentais em geral) com a realidade. Sua proposta consiste em buscar tal relação não mais nas entregas sensíveis ou num papel peculiar dado à experiência. Em termos

epistêmicos, não temos nenhuma entidade privilegiada que explique essa relação, havendo apenas um jogo interno de justificação que já acontece inteiramente dentro do domínio mental. A relação entre os habitantes da mente e o mundo objetivo não deve ser procurada na experiência sensível. Ao abandonarmos os intermediários epistêmicos, deixamos a ordem das justificações girar por conta própria. Por isso Davidson irá buscar a relação com a realidade objetiva em outro lugar e por outros meios.

### A crítica de McDowell

McDowell vê em Davidson um esforço louvável, mas acredita que ele tomou a “direção errada” (MCDOWELL, 2005, p. 138). Seu erro, para McDowell, advém de uma cegueira que poderia ser corrigida caso olhemos para a questão a partir de lentes kantianas.

O conteúdo empírico, em Davidson, opõe-se dualisticamente ao domínio conceitual. Segundo McDowell, essa oposição corresponde à distinção kantiana entre intuição (*Anschauung*) e conceito (*Begriff*). A intuição é a entrada sensória, aquilo que a faculdade da sensibilidade (*Sinnlichkeit*) oferece à mente. Nossas crenças ou teorias são significativas, ou não-vazias, por causa dessa “interação entre o conceitual e o sensorial” (MCDOWELL, 2005, p. 144).

O caráter de ser não-vazio merece atenção especial. A famosa máxima kantiana diz que “pensamentos sem conteúdo são vazios” (KANT, *KrV* A51/B75). Ao dizer isso, Kant aponta para um tipo específico de vacuidade. Um pensamento vazio, propriamente falando, não é um pensamento. Faltaria a ele aquilo que é pensado, seu conteúdo (*Inhalt*). A intuição, nessa imagética kantiana, é fornecida pela experiência, e constitui essencialmente o próprio pensamento. O jogo entre intuição e conceito responde, em Kant, a uma angústia propriamente transcendental, acerca da possibilidade mesma dos pensamentos. Não devemos pensar que a máxima kantiana aponta para a estranha classe dos pensamentos sem

conteúdo, ou vazios, mas sim para uma condição de possibilidade do pensamento.

Em termos kantianos, é preciso juntar intuição e conceito para termos um pensamento sobre o mundo. As intuições sensíveis correspondem à maneira como somos afetados pelo mundo, sendo a sensibilidade (*Sinnlichkeit*) a faculdade da mente de ser afetada. Ela é, para Kant, uma forma de receptividade (*Receptivität*). Já o entendimento (*Verstand*), outra faculdade da mente, responde pela espontaneidade (*Spontaneität*) do pensamento, que é nossa capacidade de articular ou manipular conceitos. De um lado, a intuição sensível, de outro, a faculdade de pensar acerca do que a primeira fornece. A espontaneidade do entendimento está ligada às operações envolvendo relações racionais, relações que fazemos ao pensar, julgar ou justificar algo. Nessas operações da mente, assumimos responsabilidade pelas operações, sendo nossos movimentos sensíveis a razões. O espaço racional aberto pelo exercício do pensamento constitui o que Kant chamou de reino da liberdade (*Reich der Freiheit*). Liberdade aqui deve ser entendida não como ausência de constrangimentos, mas como autodeterminação responsável ou sensibilidade a razões.

O papel desempenhado pela experiência em Kant se assemelha, num sentido importante, ao papel que ela desempenha no pensamento de Quine. Ambos retêm algo do empirismo tendo em vista a viabilização transcendental do pensamento. Sem um controle racional exercido pela experiência, seria ininteligível a objetividade do pensamento. O papel das intuições sensíveis em Kant assemelha-se ao papel desempenhado pelas impressões sensíveis em Quine<sup>7</sup>. Assim como Kant precisa da receptividade para controlar o exercício da espontaneidade, Quine precisa do tribunal da experiência para limitar a extensão da soberania conceitual do homem. Kant e Quine pretendem responder à mesma angústia transcendental. O ponto central da crítica de McDowell a Davidson é que

---

<sup>7</sup> Para uma crítica dessa comparação que McDowell faz entre Kant e Quine, ver Gibson (1999).

este seria insensível a essa motivação transcendental mais profunda, daí a direção errada que ele acabaria tomando.

Ao criticar Quine, Davidson teria apontado para um problema propriamente epistemológico. O problema, diz Davidson, é que nossas crenças não podem ser justificadas, e as evidências oferecidas são subjetivas. Quine foi visto como suscetível ao ceticismo e ao subjetivismo. Davidson não acusou Quine de adotar uma noção de experiência incapaz de fazer a devida junção entre mundo e rede conceitual, o que teria tornado o próprio pensamento sobre o mundo uma impossibilidade. Mas é essa crítica transcendental, argumenta McDowell, que Davidson deveria ter feito. Ele errou o diagnóstico e acabou seguindo um caminho equivocado.

Ao contrário de Davidson, a estratégia de McDowell não será a de eliminar da experiência seu papel racional, mas sim o contrário. Ele criticará a noção quineana de experiência como meras estimulações nervosas e tentará reabilitar uma noção kantiana de experiência (*Erfahrung*), segundo a qual é permitida a ela habitar o espaço das razões. O papel desempenhado pelo tribunal da experiência em Quine é fornecer ou possibilitar a significação empírica. Tal papel é desempenhado pelo elemento empírico que Kant atribuiu às entregas da sensibilidade, na forma de intuições sensíveis. Há, no entanto, uma importante diferença entre os dois. McDowell observa que Quine opera um truque, um movimento fraudulento, que não permite que sua visão seja aceitável. O tribunal da experiência implica a emissão de veredictos sobre nossas crenças, o que constitui um controle racional das operações do entendimento a partir da experiência. Ao mesmo tempo, Quine concebe a experiência de forma absolutamente naturalizada, como a “estimulação de receptores sensíveis”. Ou seja, McDowell concorda com Davidson quando este acusa Quine de incorrer no mito do dado. Porém, o tribunal da experiência de Quine não servia apenas para justificar nossas crenças. Ele servia, fundamentalmente, para calcar nosso pensamento no mundo. Uma vez que o tribunal se mostra inviável, não podemos simplesmente eliminá-lo, sem mais. Isso faria com que a soberania conceitual humana ficasse

descontrolada, impossibilitando o próprio pensamento. Ao defender sua perspectiva coerentista, que elimina o tribunal, Davidson deixa o pensamento operar sem qualquer constrangimento externo. Ao manter apenas relações causais com o mundo, Davidson constrói um domínio que não assume qualquer responsabilidade diante dos impactos brutos que vêm de fora das fronteiras do pensamento. Seja lá o que for que o mundo causar no interior do espaço dos conceitos, não somos responsáveis por seus reflexos. O impacto causal do mundo é uma força alienígena ao domínio racional. Podemos ser acometidos por tal impacto, mas não somos responsáveis, nem ficamos mais ou menos justificados por isso. Segundo McDowell (1994, p. 44), o dado oferece exculpações (*exculpations*) onde queríamos justificações (*justifications*). Exculpação é aquilo que explica, mas não justifica, um ato. McDowell apresenta o seguinte exemplo: alguém se encontra no lugar do qual foi banido por conta de um tornado que o conduziu até lá. O tornado explica, excuspa, o fato de ele estar lá, mas não é uma justificação para isso. O que a exculpação faz é tirar do sujeito toda a responsabilidade.

Podemos pensar em dois *desiderata*: um constrangimento racional vindo do mundo e uma espontaneidade atuante em nossa vida mental. Trata-se de mostrar como somos livres e responsáveis ao pensar, e como tal liberdade é controlada racionalmente pelo mundo. O mito do dado renuncia à liberdade do pensamento empírico, tornando-o insensível a razões. As exculpações eliminam a responsabilidade e a liberdade do nosso pensar, pois seríamos determinados a pensar da forma como pensamos. Tendo em vista esse duplo aspecto da questão, McDowell critica Davidson de ter abandonado o controle racional externo à liberdade. Ao tentar evitar o mito do dado, Davidson acabou abandonando, junto com ele, qualquer outra tentativa de encontrar um constrangimento externo à liberdade do pensamento. O coerentismo que ele irá propor responde apenas a um lado do problema. A necessidade de explicarmos a racionalidade de nossa visão de mundo nos leva, novamente, do coerentismo para a busca de alguma forma de dado. Eis o que McDowell (2005, p. 9) detectou como uma

“interminável oscilação”: do mito do dado para o coerentismo, e deste de volta para o mito.

Não irei aqui explorar a terapia que McDowell propõe para sairmos da “interminável oscilação” detectada em Davidson. No que se segue, pretendo verificar se a crítica de McDowell realmente procede, principalmente no que se refere especificamente à acusação de que ao permitir apenas relações causais entre mente e mundo, Davidson teria impossibilitado as relações racionais, tornando seu pensamento vítima da angústia transcendental que Kant, Quine e o próprio McDowell pretenderam evitar.

### **A relação entre causa e razão**

Diante da crítica de McDowell, Davidson (1999a, p. 135) afirmou certa vez que:

McDowell acha minha abordagem causal da percepção profundamente equivocada; em sua opinião, não teria explicado como a experiência provê bases racionais para a crença. Penso que o suporte mútuo entre nossas crenças perceptivas é tudo o que temos e tudo que precisamos.

O motivo de Davidson acreditar que seu coerentismo e a abordagem causal da relação da mente com o mundo não serem problemáticos em nenhum sentido transcendental mais profundo se deve, entre outras coisas, à forma como ele compreende a relação entre causa e razão. Tal sutileza, que passou despercebida por McDowell, possibilita uma maneira alternativa de compreendermos o coerentismo de Davidson, tornando-o imune às críticas apresentadas por McDowell.

Apesar da evidente defesa de uma relação causal entre mente e mundo, Davidson insiste que essa é apenas uma parte da história. Se queremos compreender o processo epistêmico de justificação, devemos isolar as crenças e admitir apenas relações causais com o exterior. Mas se queremos “uma abordagem de como muitas de nossas crenças perceptivas

vêm a ter o conteúdo que têm” (DAVIDSON, 1999b, p. 106), então precisaremos compreender o que o impacto causal pode fornecer ao sistema de crenças. As crenças perceptivas são abertas à revisão e são causalmente sensíveis ao mundo: há uma racionalidade operante tanto na criação das crenças quanto no ajuste posterior. Segundo Davidson, a fricção entre crenças e mundo, que torna o conteúdo empírico possível, está lá, nas causas de nossas crenças perceptivas. Não há necessidade de mais nada.

Em outra oportunidade, Davidson (2001, p.289) afirma que:

Tenho a impressão de que McDowell pensa que, uma vez que nossas crenças são “livres”, tal causalidade estaria fora de lugar; é um aspecto fundamental do seu dilema que o mundo das ciências naturais seja dominado pela causalidade, enquanto a vida da mente pensante não seja. Esse é um problema que não surge dessa forma se rejeitarmos tal dualismo em relação à causalidade. Pensamento, percepção, lembrança e ação voluntária caracterizam certos objetos físicos, e (na minha visão) a causalidade não “constrange” essas atividades, mas as permite.

Segundo Davidson, Kant defendia, acertadamente, que não há um conflito entre o tipo de autonomia do juízo e do pensamento e a causalidade universal. Quando falamos de eventos mentais, não podemos estabelecer leis estritas (*strict laws*), universalmente válidas e deterministas, como estabelecemos em alguns domínios das ciências naturais. Ainda assim, Davidson (2001, p. 289) acredita que “o discurso causal abarca esses domínios”.

Para Davidson, meras sensações não podem ser razões, mas uma crença perceptiva pode tomar aquilo que é percebido ao mesmo tempo como causa e razão. Ser uma razão significa simplesmente introduzir o mesmo evento que causou a crença numa explicação racional. O motivo de uma mera sensação não poder ser razão de nada é que *ser razão para algo* envolve uma relação lógica que apenas objetos conceituais podem estabelecer entre si. Mas uma causa pode ser uma razão, ou uma razão efetiva é sempre a causa de algo. Essa relação merece maiores explicações.

Em seu primeiro artigo que alcançou notoriedade, *Actions, Reasons, and Causes* (1963), Davidson defendeu que as razões, intenções ou motivos que explicam o comportamento de alguém podem também ser causas do comportamento explicado. Davidson pretendia, com sua argumentação, refutar um entendimento bastante difundido por certa tendência neowittgensteiniana<sup>8</sup>. Segundo essa tendência, uma vez aceita a concepção humeana de causalidade (segundo a qual as causas são logicamente independentes de seus efeitos), segue-se que as razões não podem ser causas de ações, pois há, entre uma coisa e outra, um elo lógico ou conceitual. Davidson viu nessa perspectiva uma estranha consequência: se as razões não podem causar ações, então elas não servem para explicar o comportamento humano efetivo.

Um primeiro passo crítico de Davidson consistiu em rever a noção de causalidade humeana. Hume está certo, segundo Davidson, ao defender que todo enunciado causal singular está comprometido com a existência de uma lei geral que ele instancia, mas erra ao pensar que razões, motivos e desejos estão nomologicamente ligados a ações. O erro está no fato de ele tomar a explicação de ações humanas através de razões como um tipo de relação regulada por leis, assim como fazemos com acontecimentos no mundo físico<sup>9</sup>.

A manobra de Davidson consiste, então, em negar a conclusão (razões não são causas), apesar de aceitar seus pressupostos (toda causa implica uma lei, e as racionalizações não implicam leis). Davidson faz isso enfraquecendo o primeiro pressuposto, pois apesar da noção humeana de causalidade exigir que cada enunciado causal particular pressuponha uma lei, ela não exige que tal lei possa ser formulada, ou tornada explícita, nos próprios termos em que a relação causal se encontra expressa. Eventos no mundo se relacionam causalmente, mas podemos descrevê-los de formas variadas. Algumas descrições enunciam uma efetiva relação causal, mas a

---

<sup>8</sup> Ver Ryle (1949) e Anscombe (1957).

<sup>9</sup> A crítica de Davidson à causalidade humeana é grandemente influenciada por Hart e Honoré (1959).

regra geral que elas instanciam pode não ser capaz de ser expressa no mesmo tipo de descrição utilizada.

O *insight* de Davidson consiste em dizer que a relação entre causa e efeito liga eventos que existem independentemente de suas descrições, sendo a relação causal indiferente à linguagem que usamos para descrevê-la. Para Davidson (1993, p. 195), “são os eventos que têm o poder de mudar as coisas, não as várias formas de descrevê-los”. Já as leis causais, ao contrário, dependem da linguagem, sendo aplicáveis a eventos enquanto descritos de uma forma particular, com certos predicados ou com o vocabulário apropriado. A instanciação de uma lei é uma relação lógica e depende de como articulamos nossos conceitos, de como situamos as coisas no espaço das razões.

Davidson (1980, p. xi) afirma que “a causa é o cimento do universo; o conceito de causa é o que mantém conjunta a nossa imagem do universo, uma imagem que, de outra forma, se desintegraria num díptico composto do mental e do físico”. Num certo sentido, as causas cortam de um lado ao outro os reinos da natureza (caracterizado por relações brutas entre as coisas do mundo) e o da razão (onde as relações são regidas por princípios normativos). É possível que dois eventos, um descrito como razão, motivo ou intenção, e outro descrito como uma ação executada, se liguem por relações causais, mas essa relação pode não instanciar nenhuma lei que poderia ser expressa tanto pelo vocabulário relativo às motivações psicológicas quanto pelo vocabulário relativo aos comportamentos humanos. É logicamente possível que alguém tenha a intenção de levantar o braço e que isso cause seu comportamento, mas não há nenhuma lei geral que explique esse fenômeno, nenhuma lei científica séria ou estrita (que permita previsões, contrafactuais etc.) que conecte intenções e ações<sup>10</sup>.

A tradição humeana inspirou uma série de tentativas de estender as explicações nomológicas à compreensão do comportamento humano<sup>11</sup>.

---

<sup>10</sup> Para uma exploração mais detalhada dessa questão, ver Nannini (1999).

<sup>11</sup> Quanto a essa perspectiva, ver, por exemplo, Hempel (1962).

Uma vez que toda explicação causal implica uma lei geral, buscou-se explicar o comportamento humano através da aplicação de modelos de predição típicos das ciências naturais. A ideia era explicar e prever o comportamento humano a partir de regularidades generalizáveis em leis. Já a tradição neo-wittgensteiniana, que também aceita o caráter nomológico da causalidade, reduziu a aplicação do discurso causal ao reino da natureza, proibindo-o nas ciências humanas e no estudo da linguagem e do pensamento. Davidson procurou uma via média: sua explicação do comportamento humano é, ao mesmo tempo, causal e não-nomológica.

A defesa de uma abordagem causal do comportamento enfrenta dois problemas distintos. Há um problema ontológico (quanto à relação entre a ação e as razões que a explicam) e um problema epistêmico (relativo ao que constitui uma explicação por razões, ou uma racionalização)<sup>12</sup>. Os neo-wittgensteinianos, ao rejeitarem teorias causais da ação, apontaram apenas para a tese epistêmica de que explicações por razões não poderiam ser explicações causais, mas esse argumento não levou em conta a tese ontológica de que razões podem ser causas. A distinção entre esses dois níveis explica a sutileza do argumento de Davidson.

Podemos falar da causa de um evento sem conhecermos, nem sermos capazes de apresentar, uma explicação causal adequada para ele (ou seja, sem termos uma explicação de tipo nomológico-dedutiva). Isso é possível uma vez que a constatação de uma relação causal particular se deve ao fato de identificarmos uma possível causa, e não somos obrigados, para tanto, a conhecermos exaustivamente todos os elementos causalmente relevantes para o efeito. No caso da explicação causal da ação, buscamos não uma teoria completa, determinista e projetável a todos os casos futuros, mas apenas a racionalização de uma ação particular. Ao explicarmos a ação de um agente, identificamos não apenas qualquer razão, mas aquela *por causa da qual* a ação foi executada. Para Davidson (1963, p. 9), “é fundamental para a relação entre uma razão e a ação que

---

<sup>12</sup> Ver Rainone (1999).

ela explica a ideia de que o agente realizou a ação porque tinha aquela razão”. Temos, nesse domínio, critérios de relevância e simplicidade que conformam os próprios conceitos que usamos. Não podemos esperar desse espaço descritivo nada que se assemelhe às previsões rigorosas e nomológicas da física.

Davidson (1967) distingue o enunciado causal (*causal statement*), caracterizado por uma semântica extensional, da explicação causal (*causal explanation*), semanticamente intensional. Ao racionalizarmos um evento e o descrevermos em termos de razões e propósitos, temos uma explicação causal que, de certa forma, toma a relação causal sob um certo aspecto, descrevendo as causas segundo um vocabulário que atende a fins específicos. No nível ontológico, onde habitam os eventos, razões são eventos descritos de uma certa forma (por exemplo, como um evento mental), e *qua* eventos que são, podem perfeitamente ser as causas de ações, que também são eventos descritos de uma forma particular. O argumento davidsoniano parte da pressuposição ontológica de que existem entidades tais como eventos. Junta-se a isso a tese de que a forma lógica das racionalizações (ou explicações por razões) é uma sentença do tipo ‘*x* causou *y*’. Por fim, adiciona-se a tese de que sentenças causais apresentam uma semântica extensional, ou seja, que um evento causa outro independentemente de como ele é descrito (podemos trocar designadores de eventos co-referentes numa sentença causal *salva veritate*).

A moral da história é que o poder explicativo das ações humanas a partir de razões depende, ao mesmo tempo, de suas relações racionais e de seu poder causal. A razão mais forte para uma ação é geralmente a que levou o agente a executá-la. É claro que existe a possibilidade de akrasia e de outras formas de irracionalidade, mas tais fenômenos não podem ser a regra, ou eliminamos a possibilidade da agência e da explicação por razões (DAVIDSON, 1971). Falar em ação é descrever certos eventos a partir de sua causa racional. Descrever algo como uma ação envolve sua inserção numa rede holística que se guia por considerações racionalizantes.

Entendemos uma ação situando-a num cenário normativo mais amplo, e relacionando-a com crenças e valores do agente, bem como com o mundo que nos cerca. Se a psicologia é essa ciência interpretativa direcionada à ação humana, então ela busca, ao mesmo tempo, o significado e as razões de cada comportamento, bem como uma eficácia causal dessa explicação. Esse último elemento da psicologia não a compromete com a existência, em seus domínios, de leis causais estritas. Crenças, razões e motivos são também pro-attitudes, inclinações ou disposições para agir de uma ou outra forma. Aliás, se não houvesse uma regularidade no comportamento, ele não seria interpretável. Como afirmou Davidson (1985, p. 245), “é um aspecto básico de uma crença ou de um desejo que ele (ou ela) irá causar certos tipos de ação em certas condições apropriadas”.

Uma outra forma de entender o ponto em questão é simplesmente constatar que a psicologia lida com um domínio regulado por princípios gerais de racionalidade. Os agentes da psicologia são livres, se guiam por razões, e se submetem, para serem o que são, a princípios normativos que explicam a própria interpretação deles como agentes. Os objetos da física, ao contrário, apenas manifestam regularidades e não obedecem a nenhum ideal normativo. Enquanto procedimento de investigação, “a psicologia é representada como uma *investigação hermenêutica*, ao invés de ser uma ciência preditiva” (KIM, 1985, p. 383). Uma explicação bem sucedida, especialmente quando aplicada a ações aparentemente obscuras ou irracionais, depende da nossa habilidade de identificar razões que se enquadrem de forma razoável e coerente num contexto mais amplo de ações e atitudes do agente. Ou seja, uma explicação racional, ao contrário de explicações puramente causais, é sensível a padrões gerais de racionalidade. Mas ao ser inserida na rede de razões, não retiramos a força causal da motivação psicológica. Nesse novo espaço, temos eventos que operam ao mesmo tempo como causa e razão de algo.

A relação causal entre mente e mundo que Davidson permite em seu coerentismo deve ser pensada da mesma forma. O fato de a relação perceptiva ser uma relação causal bruta com o mundo não impede que um

evento perceptivo seja inserido numa rede mais ampla e venha a ser interpretado como sendo uma razão para se crer em algo. Isso não viola o mito do dado, pois a causa funciona como razão não por ser o fato do mundo natural que ela é, mas por ser racionalizada de determinada forma no interior de um discurso conceitual.

A argumentação davidsoniana parte, basicamente, de três princípios: (a) princípio da interação causal (*Principle of Causal Interaction*), segundo o qual eventos (sejam eles descritos como mentais ou físicos) interagem causalmente entre si; (b) princípio do caráter nomológico da causalidade (*Principle of the Nomological Character of Causality*), que afirma que eventos relacionados causalmente instanciam uma lei geral; e (c) princípio do anomalismo do mental (*Principle of the Anomalism of the Mental*), que nega a existência de leis deterministas capazes de prever ou explicar eventos mentais (DAVIDSON, 1970, p. 208). Os três princípios, para Davidson, devem ser mantidos. Mas mantê-los, todos juntos, de forma coerente, exige as distinções apontadas nos parágrafos anteriores, que basicamente anunciam sua famosa teoria do anomalismo do mental. Tal perspectiva admite a compatibilidade desses três princípios pois a causalidade é vista como uma relação entre eventos particulares, enquanto a racionalidade remete à forma como esses eventos serão descritos. As descrições relacionam tipos (*types*) de eventos entre si, e não eventos particulares (*tokens*). Apesar de um mesmo evento isolado (*token*) poder ser descrito, ao mesmo tempo, em termos físicos e em termos mentais, não podemos estabelecer leis estritas relacionando os dois (leis psicofísicas), pois leis relacionam tipos de eventos (descritos num vocabulário específico) entre si, e não eventos isolados. A lei vale em certos domínios *descritivos*, enquanto a causalidade, que relaciona *tokens* e não *types* entre si, vale em qualquer lugar. No final das contas, podemos afirmar com Davidson que o mundo é um só, mas há várias formas de falarmos dele. A forma como falamos não altera o poder que uma coisa tem de causar outra, mas altera aquilo que podemos ver como razão do

quê. E ao situarmos uma coisa como sendo razão de outra, não a tiramos do mundo das causas, apenas lhe atribuímos um outro nível de descrição.

Davidson (1970) relaciona a anomalia do mental com a nossa liberdade (*self-rule*). O fato de o reino mental ser anômalo, ou seja, de não instanciar uma lei geral (*failure to fall under a law*), permite que exista uma esfera legítima para a liberdade humana. Seria misterioso falarmos em liberdade se nossos estados mentais, tais como crenças, desejos ou intenções, não escapassem do determinismo da natureza. Uma vez que adentramos nos domínios mentais, adentramos num reino anômalo que não submete seus habitantes às leis estritas da natureza. Davidson não pretende apenas compatibilizar autonomia e determinismo: seu objetivo mais profundo é reconciliar a liberdade mental humana com a necessidade causal do mundo. Ele busca essa façanha ao caracterizar os fenômenos mentais como metodologicamente irreduzíveis e, ao mesmo tempo, como supervenientes em relação a um nível ontológico mais básico. Entidades mentais não são, propriamente falando, uma classe diferente de seres ontologicamente independentes, mas apenas uma forma peculiar, normativamente guiada, de falarmos de certas entidades mundanas. A diferença entre níveis de descrição explica por que a liberdade é possível, mesmo para criaturas que habitam o reino da natureza.

## Conclusão

A forma como Davidson entende a relação entre causa e razão simplesmente não torna essas noções excludentes. Ser uma causa simplesmente não significa não ser uma razão, e vice e versa. Quando McDowell critica o fato de Davidson permitir *apenas* relações causais entre mente e mundo, ele não levou em consideração o fato de que a mesma coisa que causa outra pode também ser uma razão dessa outra coisa. Ou seja, se o problema transcendental exige uma relação racional com o mundo, tal relação parece ser perfeitamente permitida pelo pensamento

de Davidson. A crítica de McDowell, portanto, é injusta e não explora as várias possibilidades abertas pelo pensamento de Davidson.

Ao reconstruir o caminho que começa em Quine, passa por Sellars e chega em Davidson, McDowell percebe bem um determinado conjunto de razões que levaram Davidson a propor seu coerentismo. Mas ao exigir de Davidson uma resposta ao problema transcendental, que aparece de forma paradigmática no pensamento de Kant, McDowell acaba empobrecendo o pensamento de Davidson e não faz jus aos elementos de que ele dispõe para responder a essa questão específica. Quando olhamos com mais cuidado para a forma como Davidson entende a relação causal, vemos que ela de fato pode oferecer tudo o que precisamos para racionalizar nossas crenças. A tal angústia transcendental que McDowell detecta em Davidson surge não de uma limitação de seu pensamento, mas sim de uma leitura apressada que não problematizou adequadamente em que consiste a relação causal, e como ela pode conviver com as relações racionais.

## Referências

ALVES, Marco Aurélio Sousa. **Sobre a possibilidade de pensarmos o mundo**: o debate entre John McDowell e Donald Davidson. Dissertação de Mestrado, FAFICH/UFMG, Belo Horizonte, 2008.

ANSCOMBE, G. E. M. **Intention**. Harvard University Press, 2000 [1957].

BRANDOM, Robert B. Non-inferential knowledge, perceptual experience, and secondary qualities: placing McDowell's empiricism. In: SMITH, Nicholas H. (ed.). **Reading McDowell**: On mind and world. Routledge, 2002, p. 92-105.

DAVIDSON, Donald. Actions, Reasons, and Causes. In: \_\_\_\_\_. **Essays on actions and events**. Oxford University Press, 1980, p. 3-20 [1963].

\_\_\_\_\_. The logical form of action sentences. In: \_\_\_\_\_. **Essays on actions and events**. Oxford University Press, 1980, p. 105-21 [1967].

\_\_\_\_\_. Mental Events. In: \_\_\_\_\_. **Essays on actions and events**. Oxford University Press, 1980, p. 207-24 [1970].

\_\_\_\_\_. Agency. In: \_\_\_\_\_. **Essays on actions and events**. Oxford University Press, 1980, p. 43-62 [1971].

\_\_\_\_\_. On the Very Idea of a Conceptual Scheme. In: \_\_\_\_\_. **Inquiries into truth and interpretation**. Oxford University Press, 1984, p. 183-98 [1974].

\_\_\_\_\_. Introduction. In: \_\_\_\_\_. **Essays on actions and events**. Oxford University Press, 1980, p. xi-xvi.

\_\_\_\_\_. A coherence theory of truth and knowledge. In: \_\_\_\_\_. **Subjective, intersubjective, objective**. Oxford University Press, 2001 p. 137-53 [1983].

\_\_\_\_\_. Reply to J. J. C. Smart. In: VERMAZEN, Bruce; HINTIKKA, Merrill B. (eds.), **Essays on Davidson: actions and Events**. Oxford University Press, 1985, p. 244-7.

\_\_\_\_\_. The myth of the subjective. In: \_\_\_\_\_. **Subjective, intersubjective, objective**. Oxford University Press, 2001, p. 39-52 [1988].

\_\_\_\_\_. Thinking causes. In: \_\_\_\_\_. **Truth, language, and history**. Oxford University Press, 2005, p. 185-200 [1993].

\_\_\_\_\_. Reply to Roger F. Gibson. In: ZEGLEN, Urszula (ed.). **Donald Davidson: truth, meaning and knowledge**. Routledge, 1999a, p. 134-5.

\_\_\_\_\_. Reply to John McDowell. In: ZEGLEN, Urszula (ed.). **Donald Davidson: truth, meaning and knowledge**. Routledge, 1999b, p. 105-8.

\_\_\_\_\_. Comments on Karlovy Vary Papers. In: KOTATKO, Petr; PAGIN, Peter; SEGAL, Gabriel (eds.). **Interpreting Davidson**. Stanford: CSLI Publications, 2001, p. 285-307.

GIBSON, Roger F. McDowell on Quine, Davidson and epistemology. In: ZEGLEN, Urszula (ed.). **Donald Davidson: truth, meaning and knowledge**. Routledge, 1999, p. 123-133.

HART, H. L. A.; HONORÉ, A. M. **Causation in the law**. 2. ed. Oxford University Press, 1985 [1959].

HEMPEL, Carl G. Rational Action. In: FETZER, James H. (ed.). **The philosophy of Carl G. Hempel: studies in science, explanation, and rationality.** Oxford University Press, 2001, p. 311-27 [1962].

KANT, Immanuel. **Kritik der reinen Vernunft.** Köln: Könnemann, 1995.

\_\_\_\_\_. **Crítica da Razão Pura.** 5.ed. Trad. Manuela Pinto dos Santos e Alexandre F. Morujão. Lisboa: Calouste Gulbenkian, 2001.

KIM, Jaegwon. Psychophysical Laws. In: LEPORE, Ernest; MCLAUGHLIN, Brian (eds.). **Actions and events: perspectives on the philosophy of Donald Davidson.** Blackwell, 1985, p. 369-86.

McDOWELL, John. **Mind and World.** 2. ed. Harvard University Press, 2000 [1994].

\_\_\_\_\_. **Mente e Mundo.** Trad. João Virgílio Cuter. Aparecida, SP: Ideias & Letras, 2005.

NANNINI, Sandro. Physicalism and the anomalism of the mental. In: DE CARO, Mario (ed.), **Interpretations and causes: new perspectives on Donald Davidson's philosophy.** Dordrecht: Kluwer Academic Publishers, 1999, p. 101-16.

QUINE, Willard van Orman. Two dogmas of empiricism. In: \_\_\_\_\_. **From a logical point of view.** 2. ed. Harper & Row Publishers, 1963, p. 20-46.

\_\_\_\_\_. Epistemology naturalized. In: \_\_\_\_\_. **Ontological relativity and other essays.** New York: Columbia University Press, 1969, p. 69-90.

\_\_\_\_\_. **The roots of reference.** La Salle, Illinois: Open Court, 1974.

\_\_\_\_\_. **Word and object.** Cambridge, Mass.: The MIT Press, 1997.

RAINONE, Antonio. Thirty-five years after "Actions, reasons, and causes": what has become of Davidson's causal theory of action? In: DE CARO, Mario (ed.). **Interpretations and causes: new perspectives on Donald Davidson's philosophy.** Dordrecht: Kluwer Academic Publishers, 1999, p. 125-36.

RYLE, Gilbert. **The concept of mind.** Chicago: The University of Chicago Press, 1984 [1949].

SELLARS, Wilfrid. **Empiricism and the philosophy of mind.** Harvard University Press, 1997 [1956].



## O empirismo cognitivo de Sylvain Auroux

*Argus Romero Abreu de Moraes*<sup>1</sup>

### Considerações iniciais

O debate em torno da internalidade – de tradição racionalista – ou externalidade – de tradição empirista – da cognição humana ainda é um dos mais importantes na Linguística. De um lado, o Gerativismo continua a ser o campo de estudos mais consolidado acerca do inatismo da linguagem; do outro, distintas abordagens teóricas discordam da redução das estruturas linguísticas a módulos cerebrais humanos, tal como o Empirismo Cognitivo (AUROUX, 1992; 1993; 1998; 2009; 2010).

No presente texto, analisaremos as principais contribuições do Empirismo Cognitivo para os estudos da linguagem. Para tanto, na primeira seção, discutiremos o conceito de linguagem; na segunda, trataremos do externalismo cognitivo; na terceira, e última, apresentaremos as hipóteses das ferramentas linguísticas, da subdeterminação gramatical, da determinação histórica dos sentidos e da conjectura sociológica.

---

<sup>1</sup> Professor colaborador e pós-doutorando no Programa de Mestrado em Letras da Universidade Federal de São João del-Rei (PNPD/CAPES). E-mail: [argusromero@yahoo.com.br](mailto:argusromero@yahoo.com.br)

Como o leitor poderá notar, no intuito de garantir a fluência do texto, optamos por focar a exposição em *La raison, le langage et les normes*, de Sylvain Auroux, considerada a sua principal obra. Tomando-a por fundamento, enquadraremos seus outros textos e relações epistemológicas por notas de rodapé, destacando, em especial, o nosso interesse em realizar o diálogo entre o Empirismo Cognitivo (doravante, EC), a Análise do Discurso Francesa (doravante, AD) e a Linguística Cognitiva (doravante, LC)<sup>2</sup>. Desse modo, o texto oscila entre uma resenha crítica de Auroux (1998) e um artigo teórico sobre a relação entre pensamento, linguagem e história, na interface entre a Filosofia da Linguagem e a Filosofia da Linguística.

Para um pesquisador iniciante, as notas de rodapé podem ser excessivas, de maneira que procuramos garantir ao texto uma coerência à parte. Com isso, entendemos que a leitura pode, inclusive, prescindir dos complementos de informações. No entanto, com o fito de assegurar o rigor da interpretação, decidimos acrescentar citações e referências que pudessem, a um só tempo, fundamentar os nossos raciocínios e apresentar maior densidade teórica para um leitor mais experiente e especializado. Esperamos ter sido exitosos no cumprimento dessas propostas.

## **1. A linguagem humana como sistema real, simbólico e imaginário**

De acordo com Auroux (1998), nas *Investigações filosóficas*, Wittgenstein (1999, §185-242) desenvolve dois argumentos no intuito de demonstrar as dificuldades em se comprovar que um sujeito segue uma dada regra ao agir. O primeiro deles aponta: tomemos uma dada regra R e o comportamento de um dado sujeito que age em consonância com essa regra. Em última instância, afirma o filósofo, não seria possível considerar que esse sujeito se comportou desse modo por ter seguido tal regra. Para Wittgenstein (1999), pode-se demonstrar que existem distintas regras que dariam conta desse mesmo comportamento, uma vez que um mesmo

---

<sup>2</sup> Trabalhamos esse aspecto mais diretamente em Morais (2014; 2015b).

comportamento (inclusive o linguístico) pode estar de acordo com um conjunto indefinido de regras, impedindo-nos de demonstrar qual delas seria aquela que, de fato, orienta esse sujeito em sua ação. O segundo argumento seria uma espécie de versão introspectiva desse primeiro paradoxo. Digamos que um sujeito defina uma dada regra para produzir algo, no caso de ele ter se utilizado dela para produzir esse “algo”, não seria possível saber se a regra formulada por esse mesmo sujeito nessa ação é a causa do seu próprio comportamento<sup>3</sup>.

Consoante Aurox (1998), Kripke (1982) retomou esse raciocínio de Wittgenstein (1999)<sup>4</sup> para sustentar a impossibilidade de existência de uma linguagem privada. Segundo o estudioso francês, o filósofo austríaco desenvolveu esses argumentos para demonstrar o “hiato” incontornável entre a representação consciente que o sujeito pode ter de uma dada regra<sup>5</sup> e o processo de seguimento dessa regra na mente desse mesmo sujeito no momento em que manifesta um dado comportamento.

---

<sup>3</sup> Aurox (2009, p. 88-89. Itálicos do autor) retoma esse raciocínio do seguinte modo: “Imaginemos que Paulo queira criar uma palavra, por exemplo, ‘nefo’, para designar determinada sensação que ele experimenta um dia e decida registrá-la em seu diário toda vez que voltar a experimentá-la. Suponhamos que, alguns dias depois de ter tomado essa resolução, ele experimente uma sensação x e decide que se trata de um ‘nefo’. Sobre qual critério ele pode se apoiar para afirmar que seu uso é correto, isto é, está de acordo com sua decisão anterior? Paulo poderá responder que é porque ele se lembra de ter decidido assim. Mas talvez sua lembrança não seja exata e sua atitude provenha simplesmente do fato de que sua utilização de ‘nefo’ lhe parece correta. Não há quem possa dizer, nem mesmo o próprio Paulo, se, ao utilizar ‘nefo’, ele está de acordo com uma regra determinada ou se tem a mera impressão de estar seguindo uma regra. Enquanto apresentava seu paradoxo, Locke [*Ensaio sobre o entendimento humano*, C.XI.111] pensava que a solução consistia em construir convenções, a fim de que a linguagem não fosse ‘a posse privada do indivíduo, mas a medida comum do comércio e da significação’. Wittgenstein, que nunca se refere a Locke e que provavelmente o ignorava, parece seguir o caminho contrário. Como sugeriu Kripke [*Naming and necessity*, 1980] em um livrinho muito controverso, o objetivo do autor do *Tractatus* seria mostrar a impossibilidade de uma linguagem privada. É a comunidade que regula a linguagem e a significação. Posso perfeitamente ter sensações privadas (quando sofro, sou eu – ninguém mais – quem sofre!), mas, para representar a mim mesmo ou para exprimir esse sofrimento, disponho apenas de uma linguagem comum. Como o mostra Bouveresse, para Wittgenstein, a interioridade é um mito”.

<sup>4</sup> Sobre os contributos da filosofia de Wittgenstein para a AD, Pêcheux (2008, p. 49) destaca: “eu sublinharia o extremo interesse de uma aproximação, teórica e de procedimentos, entre as práticas de ‘análise da linguagem ordinária’ (na perspectiva anti-positivista de Wittgenstein) e as práticas de ‘leitura’ de arranjos discursivo-textuais (oriundas das abordagens estruturais)”.

<sup>5</sup> Aurox (2009, p. 81-2) define regra como o “enunciado de uma estipulação que basta seguir para obter, a partir de elementos dados, uma frase correta na língua em questão. A noção de ‘correto’ introduz um aspecto ‘normativo’. [...] Uma regra difere, portanto, das leis da natureza pelo fato de não ser obrigatória: o sujeito a segue ou não a segue, correndo o risco de não ser entendido. A regra difere ainda das leis da natureza ao admitir ‘exceções’, mesmo que o objetivo do gramático seja encontrar as regras que deixam de fora o menor número de exceções possível”.

Para Auroux (1998), não se pode negar a interioridade à qual apenas o próprio sujeito tem acesso, como nos leva a entender a “mentira” como fenômeno de linguagem. No entanto, isso não implica que não haja descontinuidade em todos os níveis cognitivos, desde o “interno” até o “externo”<sup>6</sup>. No seu dizer, a questão fundamental a ser inquirida não é sobre a existência ou não dos conhecimentos e da linguagem em si, mas o que possibilita a ilusão que as faz parecerem como instâncias, em si, apartadas do “eu” que conhece. A ilusão da existência de um “mundo dos conhecimentos e da linguagem” dessubjetivado decorre da nossa capacidade de tornar os sentidos “substâncias”. Dando-lhes materialidade, imaginamos atribuir-lhes também descrições que parecem torná-los instâncias discretas, isto é, apartadas do processo que os delimita como categorias (AUROUX, 1998)<sup>7</sup>.

No sentido inverso à abordagem representacional clássica, segundo a qual a linguagem externa representa o signo nos estados internos, Auroux (1998) defende que o ponto de partida deve ser sempre a exterioridade do próprio signo<sup>8</sup>. Nesse viés, os comportamentos humanos podem ser instrumentalizados em signos pelo modo como estes parecem se separar da realidade física na/através da/pela qual emergem. Para o estudioso, a metalinguagem ou processo autonímico seria uma propriedade fundamental da linguagem, de modo que todo elemento linguístico poder ser utilizado como seu próprio signo e ensinar sua própria representação. Em vista disso, no seu dizer, haveria a confusão comum nas Ciências da

---

<sup>6</sup> Nesse aspecto, Auroux (1998) aproxima-se das seguintes contribuições de Lévy (1993, p. 137): “O sujeito pensante também se encontra fragmentado em sua base, dissolvido pelo interior. [...] Qual a imagem que se sobressai desta dissolução do sujeito cognitivo em uma microsociedade biológica e funcional na base, e de sua imbricação em uma megassociedade povoada por homens, representações, técnicas de transmissão e de dispositivos de armazenamento, no topo? Quem pensa? Não há mais sujeito ou substância pensante, nem ‘material’, nem ‘espiritual’. O pensamento se dá em uma rede na qual neurônios, módulos cognitivos, humanos, instituições de ensino, línguas, sistemas de escrita, livros e computadores se interconectam, transformam e traduzem as representações”.

<sup>7</sup> A respeito da retomada de Kripke (1980; 1982) e de Wittgenstein (1999) no escopo do Empirismo Cognitivo, a qual desenvolvemos até este ponto do texto, cf. Auroux (1998, p. 244-5; 261-2).

<sup>8</sup> Bakhtin & Volochínov (2009, p. 31. Grifos dos autores) expressam um raciocínio semelhante para sustentar a relação constitutiva do ideológico com o signo: “Um produto ideológico faz parte de uma realidade (natural ou social) como todo corpo físico [...]; mas, ao contrário destes, ele também reflete e refrata uma outra realidade, que lhe é exterior. [...] Em outros termos, tudo que é ideológico é um *signo*”.

Linguagem de reduzir a algebrização da representação da linguagem àquela da linguagem em si. Em última instância, o sentido existe pela capacidade que os signos possuem de, a um só tempo, serem partes do mundo, pois são materiais, e de se destacarem da causalidade física – opaca – que lhes possibilita surgir.

O intervalo entre as relações causais e a emergência de formas produtoras de sentido permite ao signo produzir tanto a ritualização do comportamento simbólico quanto garantir a sua não-fixidez em nenhum lugar do mundo. Qualquer que seja o fragmento linguístico, ele é constitutiva e ontologicamente passível de se tornar outra coisa. A esse respeito, Auroux (1998, p. 295. Grifos nossos) destaca:

Eu posso escrever “A é signo de B”. Isso não é tão claro, pois B também é um signo, visto que eu o escrevi. Evidentemente, eu posso dizer que eu escrevi B para Pierre, que está sentado na minha frente. Mas, de uma parte, “Pierre” também é um signo e, de outra parte, aquilo que eu escrevi é que o signo de Pierre não é nem “Pierre”, nem B, mas A. Aquilo que eu escrevi tem por objetivo me representar como um signo que funciona como um signo; eu utilizei a linguagem para me representar e é isto que introduz a *circularidade*. O que eu queria dizer é que um símbolo só é um símbolo se ele é ligado a qualquer coisa que lhe é *exterior*<sup>9</sup>.

Como podemos notar, o que torna o símbolo um signo é a sua necessidade de apontar para o exterior de si mesmo, sendo esse exterior marcadamente histórico, produto das relações imaginárias contingentes do homem com o real percebido. Se a linguagem – como instância semiótica – fosse explicável por uma relação hermética consigo mesma, ela seria reduzida apenas aos símbolos, capazes de representação e de serem representáveis pelos sujeitos, mas incapazes de “criatividade”, de

---

<sup>9</sup> Citação em francês: “Je puis écrire « A est signe de B ». Ce n’est pas très clair, car B aussi est un signe, puisque je l’écris. Évidemment, je puis dire que j’ai écrit B, pour Pierre qui est assis en face de moi. Mais, d’une part, « Pierre » aussi est un signe et, d’autre part, ce que j’ai écrit c’est que le signe de Pierre ce n’est ni « Pierre », ni B, mais A. Ce que j’ai écrit a pour but de me représenter comment un signe fonctionne comme signe ; j’ai utilisé le langage pour me le représenter et c’est ce qui introduit la circularité. Ce que je voulais dire, c’est qu’un symbole n’est un signe que s’il est lié à quelque chose qui lui est extérieur” (AUROUX, 1998, p. 295).

“elasticidade” que pudesse transformá-los em algo que não a eles mesmos, tal como um comportamento, um movimento, um som, uma imagem etc. É da sua essência a abertura ao outro, de modo que não há como delimitar a fronteira entre o mesmo e o outro (AUROUX, 1998).

Juntamente com as instituições e sociedades do saber, Auroux (1998) avalia os diversos artefatos criados como forma de “memória artificial”, através dos quais conseguimos estocar e armazenar informações e pelos quais conseguimos ter acesso a produções sociais do conhecimento passíveis de uso. Não conseguimos ter acesso a grande parte das informações sem eles, bem como não dispomos de capacidade de memorização suficiente para armazenar essa memória externalizada distribuída.

Nesse viés, a linguagem organiza-se como um sistema<sup>10</sup> Real, Simbólico e Imaginário (Sistema RSI)<sup>11</sup>. Partindo do pressuposto de que existe apenas um mundo (hipótese do reducionismo ontológico), isto é, de que a linguagem pertence à ordem das realidades do mundo ao se apresentar como realidade material do próprio e que ela é possível pela relação existente entre o corpo humano e o mundo: a linguagem, portanto, é *real*. No entanto, a linguagem não faz parte do mundo como o faz qualquer objeto com materialidade física, dotado de arranjo físico-químico ou mesmo biológico (para o caso do corpo humano ou de outros seres vivos). Embora passível de redução à realidade material, dado que incapaz

---

<sup>10</sup> A nosso ver, tanto a defesa de uma abordagem externalista da linguagem, embora apartada da concepção de sujeito, quanto a definição da linguagem como fato social, e não biológico, estão presentes em Saussure (2006, p. 18. Itálicos nossos) “Não é a linguagem que é natural ao homem, mas a *faculdade de constituir uma língua*, vale dizer: um sistema de signos distintos correspondentes a ideias distintas. [...] A faculdade – natural ou não – de articular palavras não se exerce senão com a ajuda de *instrumento* criado e fornecido pela *coletividade*; não é, então, ilusório dizer que é a *língua* que faz a *unidade* da linguagem”.

<sup>11</sup> Vale ressaltar que essa trilogia também está presente nos estudos da psicanálise lacaniana. Segundo Elia (2007, p. 57. Grifos nossos): “O desejo, assim como o sujeito, é RSI: *real* no plano de sua causa, *simbólico* em sua articulação e *imaginário* em suas vias de realização, na medida em que ele se realiza sempre na direção dos objetos delimitados que constituem a realidade do sujeito, e que são regidos pela trama de sua fantasia [...]”. Para uma visão introdutória, consultar Elia (2007) e Žižek (2010). O imaginário é a instância fundadora do ideológico humano para a AD, o qual se constituiria nas/pelas contradições materiais e simbólicas dos grupos e classes sociais. Para uma aproximação entre Psicanálise, Linguística e Materialismo Histórico, consultar Henry (1992).

de somar algo a mais ao real, a linguagem possui uma realidade diferente, com uma nova ordem de funcionalidades (AUROUX, 1998)<sup>12</sup>.

Fora dela, não há afixos, conjugações, declinações ou complementações, como propunham os gramáticos racionalistas de Port-Royal (AUROUX, 1993, p. 63-86). Ao contrário, a linguagem define as categorias da própria linguagem. Essa relação de circularidade a põe como ponto de partida e de chegada e nos permite fazê-la funcionar como sistema composto por categorias representáveis, as quais demonstram a originalidade desse novo sistema, ancorado no *simbólico* (AUROUX, 1998).

Para o Empirismo Cognitivo, não existiria o pensamento em si, mas o pensável. Se o mundo existe sem o pensamento, o pensamento não existe sem o mundo<sup>13</sup>. Além de real e simbólica, a linguagem é partilhada, isto é, *imaginária*. O imaginário emerge da relação entre os diferentes sujeitos e o mundo, tornando-os aptos a partilhar sistemas simbólicos não absolutamente interpretáveis, capazes de conformar realidades como a do pensamento. Se a linguagem computacional, por exemplo, é real e simbólica, ela não é, contudo, imaginária. O sentido não surge, portanto, de representações de relações causais e herméticas, na relação de si consigo mesmo. É da externalidade que surge a possibilidade de interpretá-lo (AUROUX, 1998).

A mera composição de estruturas mais complexas a partir de unidades discretas mínimas – hipótese da composicionalidade – não garante a produção semântica, visto que o elemento linguístico, mais do que um

---

<sup>12</sup> A esse respeito, Auroux (2009, p. 115) sustenta: “É razoável dizer que a biologia não supõe a existência de outras entidades além daquelas analisáveis pela física e pela química. Mas isso não quer dizer que os elementos próprios da biologia (os hormônios, as leis da hereditariedade, o código genético, os neurônios etc.) sejam dedutíveis das leis da física; para relacioná-los, contamos essencialmente com os processos de evolução. Não há motivo algum pelo qual as leis da biologia impliquem as da psicologia e estas últimas leis impliquem as leis da linguagem. O reducionismo ontológico – aparentemente bem fundamentado – não implica o reducionismo funcional”.

<sup>13</sup> O real do mundo em relação ao pensamento humano está posto da seguinte forma em Althusser (1978, p. 155. Grifos do autor), uma das principais influências filosóficas para a AD: “Se o processo de conhecimento não transforma o objeto real, mas somente sua intuição em conceitos, depois concreto-de-pensamento, e se todo esse processo se passa, como repisa Marx, ‘no pensamento’, e não no objeto real, é portanto em função do objeto real, e para conhecê-lo, que o ‘pensamento’ trabalha sobre *uma outra matéria* que não é o objeto real. Ele trabalha sobre as formas transitórias que o designam no processo de transformação, para produzir finalmente seu conceito: o concreto-de-pensamento”.

símbolo, é um signo<sup>14</sup>. Logo, ele deve ser signo de qualquer coisa que não ele mesmo. Essa “coisa” é o que o obriga a apontar para o seu exterior para que possa adquirir sentido sem sair da linguagem, caracterizando a capacidade autonímica do signo<sup>15</sup>. Como nos ensina Auroux (1998), qualquer tentativa de descrever a competência linguageira dos homens desconsiderando uma dessas três instâncias estará fadada ao fracasso.

Não se nega, evidentemente, a capacidade biológica humana para a aquisição da linguagem. Fosse esse o caso, não seria possível nem o pensamento, nem a comunicação. A nossa competência para representar nos/pelos sistemas simbólicos pressupõe o processo de aprendizagem, embora este não se restrinja à prescrição genética de regras algorítmicas associadas à morfossintaxe das línguas naturais, bem como à estocagem abstrata de conhecimentos. “A linguagem humana possui incontestavelmente um fundamento genético. Ele é necessário, mas provavelmente não é suficiente”, isso significa que “nossa linguagem se nos apresenta como fenômeno histórico e contingente; não é insensato pensar que essa contingência é irreduzível”, destaca Auroux (2009, p. 22). Desenvolveremos melhor esses aspectos na seção seguinte.

## 2. Não há interior sem interiorização: o externalismo cognitivo

Auroux (1998) avança em alguns aspectos marcantes para uma teoria externalista das línguas humanas, entendendo por externalismo, é importante que fique claro, a abordagem dos estudos da linguagem que

---

<sup>14</sup> Para Auroux (2009, p. 112. Itálicos do autor), “[o] ponto essencial da formalização das línguas naturais repousa sobre a noção de cálculo. Significação ou boa formação de uma frase devem ser uma função de seus constituintes; é o que se chama de *composicionalidade*. Não é completamente evidente que uma ‘frase’ em situação disponha de uma significação determinada por força de seus componentes apenas, sem necessidade de um contexto, eventualmente não linguístico. Nessas condições, o papel de uma percepção global da situação seria mais determinante que um cálculo passo a passo”.

<sup>15</sup> Sobre a relação entre mundo e linguagem, Auroux (2009, p. 64-5) esclarece que “[e]ntender uma língua não é apenas ter acesso a relações intralinguísticas, é mais: é ser capaz de relacionar esses signos linguísticos a uma experiência compartilhada de mundo. Essa irreduzibilidade da referência corresponde a uma externalidade essencial, pela qual é o próprio mundo quem participa da significação linguística. [...] O mundo participa da significação da linguagem, que muda se a significação for transplantada de ambiente. Uma relação direta com a realidade é indispensável para a significação linguística. Eis porque, no estágio atual de nossas tecnologias, os computadores que manipulam sistemas de signos de modo intrínseco não podem ‘aprender mesmo uma língua’”.

considera as línguas naturais como artefatos históricos produzidos, adquiridos e distribuídos coletivamente pela interação constante entre indivíduos humanos e instrumentos não-humanos capazes de armazenamento desses saberes.

Nesse sentido, sua perspectiva opõe-se diretamente à do inatismo chomskyano, a qual entende que o desenvolvimento da linguagem humana pressupõe a existência de um módulo cerebral destinado especificamente a esse fim, propondo que qualquer perspectiva não biológica da linguagem deve ser interpretada como comportamentalista<sup>16</sup>. Logo, se, para o cientista da linguagem francês, a língua é uma estrutura *completamente* produzida na externalidade social, para o segundo, ela deriva *integralmente* da internalidade orgânica, sendo, neste último caso, qualquer aspecto vinculado ao uso compreendido como secundário em relação ao “núcleo duro” de computação – herdado filogeneticamente – sintática e semântica.

Para Aurox (1998), como sistema RSI, as línguas naturais diferenciam-se das artificiais. Para o estudioso da linguagem, o desenvolvimento de estruturas cognitivas externas foi potencializado pelo que intitula de “tecnologia intelectual da escrita”, a qual depende tanto do desenvolvimento de instrumentos externos – como livros, bibliotecas,

---

<sup>16</sup> Chomsky (2018 [2016], p. 31-3. Itálicos do autor) define como comportamentalista (*behaviorista*) qualquer modelo teórico que defina a língua como um sistema de hábitos e convenções de uma comunidade de fala, conforme esclarecido na síntese a seguir: “[...] Para Ferdinand de Saussure, por exemplo, a linguagem (em seu sentido relevante) é um depósito de imagens de palavras na mente dos membros de uma comunidade; ‘ela não existe senão em virtude de uma espécie de contrato estabelecido entre os membros de uma comunidade’. Para Leonard Bloomfield, a linguagem é uma série de hábitos usados para responder a situações com sons de fala convencionais e para responder a esses sons com ações. Em outro lugar, Bloomfield definiu a linguagem como ‘a totalidade dos enunciados feitos em uma comunidade de fala’ – algo como a concepção anterior de linguagem estabelecida por William Dwight Whitney: a linguagem era ‘o corpo de sinais proferidos e audíveis pelos quais, na sociedade humana, o pensamento é, via de regra, expresso’. Portanto, ‘sinais audíveis para o pensamento’ [...]. Edward Sapir definiu a linguagem como ‘um modo puramente humano e não instintivo de comunicar ideias, emoções e desejos por meio de um sistema de símbolos produzidos voluntariamente’. Com tais concepções sobre a linguagem, parece natural seguir o que Martin Joos chamou de ‘tradição boasiana’ (de Franz Boas), aceitando que a linguagem pode diferir de maneira arbitrária e que cada nova língua e que cada nova língua deve ser estudada sem preconceções. Assim, a teoria linguística consiste de procedimentos analíticos para reduzir um *corpus* a uma forma organizada; basicamente técnicas de segmentação e classificação. O desenvolvimento mais sofisticado dessa concepção foi o *Método* de Zellig Harris. Uma versão contemporânea disso é entender a teoria linguística como um sistema de métodos para o processamento de expressões”.

artefatos de cálculo e de observação – quanto de estruturas sociais de produção e acumulação de conhecimentos, compostas por enciclopédias, sociedades do saber e redes de produção e reprodução culturais. O pesquisador francês sustenta que, assim como a produção de riquezas, o processo cognitivo depende da estruturação social das relações.

Tomemos como exemplo, novamente, o caso dos computadores. Eles possuem categorias simbólicas discretas que admitem a manutenção de padrões de organização internos e coerentes no enunciado produzido (*output*). Todavia, a diferença principal entre a linguagem computacional e a humana reside no fato de esta ser produzida por uma relação com o seu exterior, isto é, com o mundo social. A linguagem artificial, ao contrário, tende à fixidez da recursividade às mesmas regras/axiomas/algoritmos de base, não adquirindo variação por ser um sistema fechado e incapaz de ser alimentado por *inputs* que não estejam previstos de antemão para a conformação dos seus enunciados. No dizer de Auroux (1998, p. 7):

O homem é antes de tudo uma estrutura biológica (um corpo) que interage com um ambiente e outros corpos. O fenômeno fundamental é a constituição de ferramentas, o que implica a instrumentalização tanto do ambiente como do próprio corpo. É bastante equivocado conceder demais à interioridade. É provável que a capacidade de calcular não tenha começado simplesmente na cabeça das pessoas e que ela não teria nascido sem a manipulação de objetos externos (seixos, ábacos etc.)<sup>17</sup>.

Nesse viés, a inteligência humana constitui-se como construção coletiva não inata e suas manifestações dependem de instrumentos externos. A inteligência seria, por conseguinte, uma estrutura representacional do mundo, da qual os indivíduos detêm apenas uma parcela. Se nem tudo está alocado na mente humana, ela, todavia, é capaz

---

<sup>17</sup> Citação no original: “L’homme est avant tout une structure biologique (un corps) qui interagit avec un environnement et d’autres corps. Le phénomène fondamental est la constitution d’outils, ce qui implique l’instrumentalisation autant de l’environnement que du corps propre. On a tort de trop concéder à l’intériorité. Il est vraisemblable que la capacité de calculer n’a pas commencé simplement dans la tête des gens et qu’elle ne serait pas née sans des manipulations d’objets externes (cailloux, abaques, etc.)” (AUROUX, 1998, p. 7).

de armazenar informações, procedimentos e sistemas simbólicos, como são os casos do cálculo matemático e das línguas naturais (AUROUX, 1998).

Parte dessas estruturas cognitivas estaria distribuída em diferentes tecnologias, as quais são (re)utilizadas nas interações cotidianas entre os sujeitos e o ambiente. Em oposição às abordagens substancialistas – nativistas – da linguagem, as quais ancoram-se no pressuposto de que a linguagem seria uma herança evolutiva da espécie humana, a história das ideias linguísticas, com foco na gramatização dos saberes linguísticos, aponta para o fato de que as regras, assim como quaisquer outros conceitos nos sistemas simbólicos humanos, são internalizadas e seguidas empiricamente. Nessa ótica, os dispositivos artificiais modificam a própria ecologia da comunicação humana (AUROUX, 1998)<sup>18</sup>.

O fato de os sistemas simbólicos possuírem estruturas próprias não subverte a possibilidade de avaliar a linguagem como fruto de um processo histórico “adquirido/interiorizado” pelas experiências sociais dos indivíduos, os quais, sendo aptos a representar, são capazes também de pensar/imaginar. A singularidade dos sujeitos decorre de suas experiências idiossincráticas e da não repetitividade de cada situação enunciativa na qual é “chamado” a enunciar. Consoante Auroux (1998, p. 297):

Não há nenhuma razão para se pensar que a linguagem não nasce na instrumentalização do próprio corpo e do ambiente. Ainda que a emergência

---

<sup>18</sup> Auroux (1992, p. 8-9. Itálicos do autor) dedica-se ao desenvolvimento de duas teses centrais ao seu pensamento teórico, a saber: “A primeira, concerne ao nascimento das ciências da linguagem (ou, para evitar o bloqueio da mitologia comparatista, digamos ‘as considerações reflexivas sobre a linguagem humana’). Os historiadores, os linguistas e os filósofos habitualmente fazem desse aparecimento uma das causas do nascimento da escrita. Em meu primeiro capítulo sustento o contrário, a escrita é que é um dos fatores necessários ao aparecimento das ciências da linguagem, as quais remontam à virada do terceiro e do segundo milênios antes de nossa era, entre os acadianos. [...] A segunda tese concerne ao que chamo de *gramatização* e é objeto do segundo e terceiro capítulos. Podemos formulá-la assim: o Renascimento europeu é o ponto de inflexão de um processo que conduz a produzir dicionários e gramáticas de todas as línguas (e não somente dos vernáculos europeus) na base da tradição greco-latina. Esse processo de ‘gramatização’ mudou profundamente a ecologia da comunicação humana e deu ao Ocidente um meio de conhecimento/dominação sobre outras culturas do planeta. Trata-se propriamente de uma revolução tecnológica que não hesito em considerar tão importante para a história da humanidade quanto a revolução agrária do Neolítico ou a Revolução Industrial do século XIX”.

da linguagem humana suponha sistemas simbólicos possuindo estruturas próprias, nada obriga a pensar que devamos imaginar que a realização dessas estruturas depende de uma representação prévia na interioridade. A possibilidade do pensamento simbólico provém mais provavelmente da capacidade de se representar essas estruturas. Não há interioridade sem interiorização. A tese última do externalismo é que o próprio pensamento é de essência histórica e empírica<sup>19</sup>.

Um sujeito com tendências ao singular não é o mesmo que um sujeito reduzido ao indivíduo. Essa não determinação é fundamental para o entendimento de que a linguagem é histórica. Se o primeiro é uma posição emergente em uma dada situação de interação, tendo por funcionamento uma causalidade imanente à própria linguagem, o segundo configura-se como o ser orgânico/biológico formado por parte das capacidades cognitivas oriundas da história evolutiva humana. Se a relação entre ambos parece indiscutível, a redução de um ao outro apresenta-se como problemática<sup>20</sup>. Dito isso, passemos às quatro principais hipóteses externalistas do Empirismo Cognitivo.

### 3. Quatro hipóteses do empirismo cognitivo

Segundo Aurox (1998), a perspectiva externalista é irreconciliável com as hipóteses da “língua” e do “cálculo”<sup>21</sup>. No caso da primeira,

---

19 Passagem no original: “Il n’y a aucune raison de penser que le langage ne naît pas dans l’instrumentalisation du corps propre et de l’environnement. Quand bien même l’émergence du langage humain suppose des systèmes symboliques possédant des structures propres, rien n’impose de penser qu’il faille imaginer que la réalisation de ces structures dépende d’une représentation préalable dans l’intériorité. La possibilité de la pensée symbolique provient plus probablement de la capacité de se représenter ces structures. Il n’y a pas d’intériorité sans intériorisation. La thèse ultime de l’externalisme, c’est que l’esprit lui-même est d’essence historique et empirique” (AUROUX, 1998, p. 297).

<sup>20</sup> De forma semelhante, Bakhtin e Volochinov (2009, p. 59) diferenciam o “indivíduo” (natural e isolado) da “individualidade” (superestrutura ideológica semiótica). Para um debate sobre essa dualidade no escopo teórico da AD, distinguindo o “indivíduo” do “sujeito”, consultar Moraes (2017). Para uma análise comparativa do conceito de natureza humana em Piaget, Chomsky e Foucault, vide Moraes (2019).

<sup>21</sup> Sobre a redução da língua à ideia de gramática, Aurox (2010, 245) esclarece: “[...] Um dos impasses da filosofia analítica da linguagem provém, a meu ver, da tentativa de construir uma teoria da referência dentro de um contexto totalmente abstrato, no nível em que estaria aquilo que nós chamamos de língua gramatical”. Sobre a hipótese da língua como cálculo, já discutido na nota 14, Aurox (2010, p. 247-8. Itálicos nossos) sustenta: “O fato de falar não é puramente e simplesmente calcular, mas se situar numa hiperlíngua, explica determinadas dificuldades na componencialidade [composicionalidade] do sentido. Na medida em que se defende essa hipótese, supõem-se duas coisas. A primeira é que se é capaz de construir uma representação semântica de cada uma das unidades linguísticas

considera-se a possibilidade de existência de uma língua interna a todos os sujeitos que seja única e homogênea, como proposto pela competência chomskyana. Isso significaria que a língua é composta por um conjunto de representações comuns/universais aos indivíduos. No tocante à segunda, o sistema linguístico seria definido como um conjunto consistente de axiomas que possibilitam descrever, explicar e engendrar as propriedades de todas as frases possíveis em uma dada língua, expressando um grupo restrito de regras capazes de predizer o conjunto possível das frases bem formadas de acordo com essa mesma língua ( $L_i$ )<sup>22</sup>.

Desse modo, o conhecimento de uma representação coletiva em uma dada ocorrência linguística poderia se efetuar pela competência linguística de apenas um indivíduo, reduzindo uma representação linguística coletiva à sua hereditariedade biológica. Privilegiando o aspecto histórico das estruturas e saberes linguísticos, Auroux (1998) considera que a organização dos Estados Nacionais catalisou a manifestação empírica da “hipótese da língua”, uma vez que a unificação de extensos territórios, englobando uma pluralidade de povos sob as mesmas “bandeiras”, funcionou como um dos principais fatores para a generalização da comunicação.

A introdução da escolaridade, a produção de gramáticas normativas – com o fito de estabelecer os “bons” e os “maus” usos das línguas – e a

---

de tal forma que o valor semântico de toda combinação dessas unidades é uma função de valor se [sic] [de] seus componentes. A segunda, que deriva da primeira, é que a representação semântica prediz verdadeiramente o valor de toda ocorrência possível. É esta restrição que impede que se possa compreender o que quer que seja da *criatividade*”.

<sup>22</sup> Sobre o nativismo – gerativismo – chomskyano, Auroux (2009, p. 51) assevera: “De certo modo, o nativismo é uma retomada trivial do transcendentalismo: imaginamos que há uma estrutura genética do espírito humano na origem das formas gramaticais. A teoria chomskyana também estava baseada na inovação técnica das gramáticas formais. Uma gramática formal é constituída de um vocabulário de base (as categorias), de axiomas e de regras de derivação. Aplicada a uma língua determinada, acredita-se que ela possa gerar todas as frases gramaticais dessa língua. [...] A relação entre a gramática universal (GU) e a gramática particular (GP) pode ser considerada de dois modos: ou a GU serve para controlar a boa formação da GP, ou a GP completa a GU com axiomas e/ou regras de derivação suplementares. Nos dois casos, fica por determinar a parte que realmente cabe à GU e o que o arbitrário da linguagem pode significar. Acrescentemos que, mesmo sendo incontestável que a linguagem – assim como todas as atividades humanas – tem um substrato biológico, a diversidade dos elementos desse substrato e de sua interação com o meio ambiente parece tamanha que torna realmente improvável que uma estrutura genética específica e única seja responsável por ela”.

criação de dicionários vernáculos aumentaram significativamente o número de dispositivos capazes de funcionar como memória externa. Segundo Auroux (1998, p. 166), esses instrumentos podem ser definidos como “uma instância material que serve de mediação durante as atividades cognitivas entre o(s) sujeito(s) do conhecimento e seu(s) objeto(s), conhecido(s) como uma realidade ou realidades do mundo”<sup>23</sup>. Sobre a instrumentalização da linguagem, Auroux (1998, p. 265) sustenta:

HIPÓTESE DAS FERRAMENTAS LINGUÍSTICAS: Uma gramática, um dicionário e os outros objetos desse tipo, tais como nós os encontramos, em particular, na sociedade ocidental, são objetos técnicos que prolongam o comportamento “natural” humano e o transformam de maneira análoga ao que produzem os objetos técnicos materiais comuns.

COROLÁRIO: As práticas languageiras observáveis não são as mesmas, haja vista que em uma dada cultura existem ou não ferramentas linguísticas; elas variam igualmente em função da natureza dessas ferramentas<sup>24</sup>.

Assim, a subdeterminação gramatical impede que um sistema de regras seja capaz de engendrar todo o conjunto de enunciados emitidos pelos locutores de uma dada língua, bem como as ferramentas linguísticas são instrumentos artificiais capazes de modificar a organização cognitiva e comunicativa humanas. Isso nos leva à compreensão de que a definição de língua como um sistema de regras rígido está associada às tendências homogeneizantes dos usos linguísticos das ordenações políticas, históricas, geográficas, jurídicas e econômicas dos Estados-Nações. Essas forças reguladoras dos usos linguísticos acarretaram o aumento extensional da semelhança entre as interações languageiras e as regras que

---

23 Citação original : “Un instrument est une instance matérielle qui sert de médiation lors des activité(s) cognitives entre le(s) sujet(s) connaissant et son (ses) objet(s), conçu(s) comme une (des) réalité(s) du monde” (AUROUX, 1998, p. 166).

24 Citação original: “HYPOTHÈSE DES OUTILS LINGUISTIQUES: Une grammaire, un dictionnaire, et les autres objets de ce type, tels qu’on les rencontre, en particulier, dans la société occidentale, sont des objets techniques qui prolongent le comportement « naturel » humain et le transforment de manière analogue à ce que produisent les objets techniques matériels.

COROLLAIRE : Les pratiques langagières observables ne sont pas les mêmes selon que dans une culture donnée existent ou non des outils linguistiques ; elles varient également en fonction de la nature de ces outils” (AUROUX, 1998, p. 265).

regem a boa formação das frases em uma dada língua. As línguas vernáculas surgem, portanto, da unificação idealizada da língua pelos Estados nos pressupostos da homogeneidade e da unicidade (AUROUX, 1998).

Não obstante, o aumento das semelhanças entre as frases proferidas pelos falantes em uma dada “língua” – racionalização da comunicação – não significa que elas façam parte da essência dessa mesma língua. Para Auroux (1998), Chomsky teria cometido o equívoco de reduzir estruturas gerais a estruturas inatas<sup>25</sup>. Em outras palavras, a existência de estruturas que possam ser utilizadas por todos os humanos não implica que elas estejam presentes nos seus genes e alocadas em seus cérebros<sup>26</sup>.

Ainda nesse viés, outro equívoco do Gerativismo seria o de ter confundido retrodição com predição. Por exemplo, o fato de ser possível “reconstituir” os percursos de uma palavra do português do século XXI em direção ao latim antigo não garante que o contrário seja possível, isto é, reconstituir na época do latim antigo as possibilidades de se chegar a uma palavra do português atual. Retroagir nos possibilita, em alguma medida, reconstituir os eventos históricos – não predizíveis – envolvidos no processo diacrônico de formação de uma dada palavra, pois dispomos de dados empíricos que ajudam nesse intento. O contrário não é verdadeiro, pois, caso estivéssemos na época do latim antigo, seria impossível prever os usos históricos futuros desse léxico (AUROUX, 1998).

Essa dissimetria impossibilita o reducionismo funcional da língua em relação ao próprio sistema e às representações mentais inatas, apontando

---

25 Essa crítica é similar à feita por Piaget a Chomsky e a Jung. Em Piatelli-Palmarini (1979, p. 231. Tradução nossa), Piaget destaca: “Não se deve confundir geral e inato; toda a fraqueza da obra de Jung é a de ter acreditado que porque um mito é geral, ele corresponde ao inato. Ora, 1) a generalidade apenas não confirma o inato, e 2) a questão é saber se a generalidade é comum a todos os níveis de desenvolvimento ou então se ela se torna geral. Em numerosos casos, as estruturas se tornam gerais, mas elas não são anteriores ao nível considerado. Citação em francês: “il ne faut pas confondre général et inné; toute la faiblesse de l’oeuvre de Jung, c’est d’avoir cru que parce qu’un mythe est général il correspond à un archétype inné. Or, 1) la généralité seule ne confirme pas l’inné, et 2) la question est de savoir si la généralité est commune à tous les niveaux de développement ou bien si elle devient générale. Dans de nombreux cas les structures deviennent générales, mais elles ne le sont pas avant le niveau considéré [...]”.

<sup>26</sup> Sobre a universalidade possível da linguagem, Auroux (2009, p. 44) sustenta que “uma língua natural é uma língua na qual todas as outras são traduzíveis; da universalidade, ela não possui unicidade, mas certamente possui a extensão, ao menos a extensão potencial”.

para o fato de que a língua, como fenômeno empírico, não possui existência autônoma. Isso explicaria, por exemplo, porque uma máquina, como a de Turing, embora consiga gerar infinitas frases bem formadas em um sistema representacional artificial dotado de um número restrito de regras/algoritmos/axiomas, não consegue pensar. Faltaria a ela um dispositivo exclusivo das línguas naturais: o imaginário (AUROUX, 1998)<sup>27</sup>.

As atividades concretas dos usuários das línguas não se reduzem a uma prescrição normativa típica de uma gramática, seja ela a tradicional/extensional ou a gramática universal/intensional chomskyana. Todos os dias, inúmeras frases ou léxicos são materializados subvertendo as variadas regras gramaticais das línguas naturais. No entanto, a incorporação ou não ao léxico ou à sintaxe dessa língua dependerá da autorização das ferramentas linguísticas, as quais são responsáveis por permitir que uma dada variação no uso possa transformar-se em regra (AUROUX, 1998).

Isso nos levaria a uma segunda dissimetria na língua: a interpretação não parece ser um processo necessariamente reversível com relação à produção. No dia a dia, é possível que nos utilizemos de expressões das quais não temos certeza sobre a sua significação. Podemos tê-las aprendido em uma aula, em um filme ou mesmo em uma interação corriqueira com alguém. Isso não nos impede de “memorizá-la globalmente” e de utilizá-la em outros contextos (AUROUX, 1998).

Durante um período, essa expressão poderá funcionar como uma espécie de “caixa-preta”, dado que a incorporamos sem ter certeza dos seus usos, até o momento em que sejamos ou corrigidos por alguém que “dominaria o bom uso” do termo ou que passemos a analisá-la de acordo com as ferramentas linguísticas que regem a sua significação (dicionários)

---

<sup>27</sup> A esse respeito, Auroux (2010, p. 249) assevera: “Um homem que não estivesse situado na realidade de uma hiperlíngua nãoalaria mais do que um computador incapaz de responder ao teste de Turing”. Sobre o teste de Turing (AUROUX, 2010, p. 249), explica: “Para testar a hipótese de que um computador é dotado da mesma capacidade linguística do homem, Turing propunha que o submetesse a uma ‘conversa’ com um ser humano que não sabia com que interlocutor ele estaria conversando: o teste seria concludente somente caso o sujeito não percebesse que ele não estava falando com um de seus semelhantes”.

e o seu uso “adequado” (gramáticas). Esse processo compõe o que Auroux (1998) chama de “controle social (consciente) das atividades linguageiras humanas”:

Na medida em que a diversidade é o modo de existência das línguas humanas, a capacidade linguística dos homens é raramente observável do exterior e sem mediação. [...] As capacidades linguísticas próprias a uma dada língua podem constituir um observatório para uma outra língua. [...] Nas trocas linguageiras, os fatos da língua ocorrem de forma dispersa e desigual. Suas homogeneizações e reuniões – que necessitam sempre de tempo e de investimentos – efetuam-se nas gramáticas e dicionários, que se tornam *ipso facto* observatórios (AUROUX, 1998, p. 168).<sup>28</sup>

Para Auroux (1998), essa opacidade diminuiria à medida que um sujeito conseguisse experienciar os usos enunciativos em uma dada prática linguageira, sem jamais, contudo, dominar completamente as relações que o sustentam (hiper)linguisticamente<sup>29</sup>. Ademais, o autor aproxima-se da perspectiva institucional da linguagem de Guilhamou (1989), de modo que as ferramentas linguísticas acabaram institucionalizando um saber metalinguístico que tem por função normatizar a comunicação humana pelo estabelecimento de regras gramaticais.

---

<sup>28</sup> Citação original: “Dans la mesure où la diversité est le mode d’existence des langues humaines, la capacité linguistique des hommes est rarement observable de l’extérieur et sans médiation. [...] Les capacités linguistiques propres à une langue donnée peuvent constituer un observatoire pour une autre langue. [...] Dans les échanges langagiers les faits de langue se donnent de façon dispersée et disparate. Leur homogénéisation et leur rassemblement – qui demandent toujours du temps et des investissements – s’effectuent dans les grammaires et les dictionnaires, qui deviennent *ipso facto* des observatoires” (AUROUX, 1998, p. 168).

<sup>29</sup> Auroux (2010, p. 243) destaca que “[e]ssa imagem me foi sugerida por Lévy (*Les Technologies de l’intelligence*, Paris, La Découverte, 1990 [1993], p. 80-2): ‘Para que as coletividades compartilhem do sentido, não é bastante que cada um de seus membros receba a mesma mensagem. O papel das coletividades é precisamente o de colocar em comum não somente textos, mas as redes de associações, de anotações e de comentários no interior das quais eles são apreendidos uns em relação aos outros. Dessa maneira, encontra-se ao dispor e materializada a constituição do sentido comum: a elaboração coletiva de um hipertexto (p. 81)’. Auroux (2010, p. 243) afirma, ainda, que “[...] o espaço-tempo, em relação à intercomunicação humana, não é vazio, ele dispõe de uma certa estrutura, conferida pelos objetos e pelos sujeitos que o ocupam. Denominaremos ‘hiperlíngua’ a esse espaço-tempo estruturado”. Por fim, Auroux (2010, p. 249) destaca que a hipótese empirista da hiperlíngua “[...] se apoia (entre outras coisas) no fato de que certas ocorrências são incompreensíveis se não se considera que elas têm, notadamente, por significação a remissão a outras ocorrências, isto é, à representação de outras ocorrências, suscetível de possuir diferentes formas (o que disse X em tal circunstância, o que eu li em tal autor etc.)”.

O essencial das práticas humanas parece ser a tendência para que se busque transformar e melhorar os mecanismos mais fundamentais e espontâneos da sua existência, de modo que a técnica passa a ser a própria essência da razão. Nessa acepção, inclusive, não devemos temer a automatização da comunicação humana, visto que ela é condizente com a característica culturalmente transformadora das nossas sociedades por meio da criação de novos instrumentos. A invenção de novos artefatos não põe a própria história em xeque, sendo, antes, parte constituinte dela (AUROUX, 1998).

Destarte, todos os sistemas linguísticos são subdeterminados gramaticalmente, na medida em que não haveria nenhum sistema simbólico passível de dar conta de todas as produções languageiras de uma dada língua. Sendo a “língua” uma criação histórica, logo, sujeita às variações de uso por grupos e pela temporalidade, nenhuma regra gramatical é capaz de encerrar todas as suas possibilidades de uso *aprioristicamente*. A criatividade, portanto, não estaria alocada na “natureza” do ser, mas na própria história, a qual tornaria possível os usos da linguagem pelos diferentes sujeitos e as estabilizações possíveis que, porventura, venham a se tornar normas através do saber gramatical. Assim, um ato do discurso, qualquer que seja, só é passível de ser interpretado e de obter eficiência comunicativa quando avaliado no interior de uma sequência histórica contingente (AUROUX, 1998)<sup>30</sup>.

---

<sup>30</sup> Importante notar a proximidade entre a análise histórica da linguagem de Guilhamou (1989) e aquela desenvolvida por Foucault (2008) a respeito do controle social dos enunciados, sendo que este último autor privilegia a relação entre subjetividade, institucionalidade e poder. Há, com isso, uma aproximação possível do Empirismo Cognitivo de Auroux (1998) com o escopo teórico da AD. Foucault (2008), por exemplo, destaca um conjunto de procedimentos externos de exclusão do dizer, tais como a interdição (p. 9-10), a oposição razão-loucura (p. 10-13), a oposição verdadeiro-falso (p. 13-9), além de procedimentos internos, quais sejam, o comentário (p. 21-6), o autor (p. 26-9) e as disciplinas (p. 29-36). Por fim, há um terceiro grupo, vinculado à ritualidade da fala, capaz de “determinar as condições de seu [do discurso] funcionamento, de impor aos indivíduos certo número de regras e assim de não permitir que todo mundo tenha acesso a eles” (FOUCAULT, 2008, p. 36). Pêcheux (2008, p. 33), por sua vez, retoma o conceito kantiano de pragmática na *Crítica da Razão Pura*, como “lei prática que tem como motivo a felicidade”, para asseverar que: “O sujeito pragmático – isto é, cada um de nós, ou ‘simples particulares’ face às diversas urgências de sua vida – tem por si mesmo uma imperiosa necessidade de homogeneidade lógica: isto se marca pela existência dessa multiplicidade de pequenos sistemas lógicos portáteis que vão da gestão cotidiana da existência (por exemplo, em nossa civilização, o porta-notas, as chaves, a agenda, os papéis, etc.) até as ‘grandes decisões’ da vida social e afetiva (eu decido fazer isto e não aquilo, de responder a X e não a Y, etc...) passando por todo o contexto sócio-técnico dos ‘aparelhos domésticos (isto é, a série de objetos que adquirimos e que aprendemos a fazer funcionar, que jogamos e perdemos, que quebramos, que consertamos e substituímos)...” (PÊCHEUX, 2008, p. 63).

Nesse sentido, Auroux (1998) defende duas hipóteses, quais sejam: a hipótese da história, a qual supõe que a temporalidade é criativa, e a conjectura sociológica, que defende que os processos cognitivos não estão presentes nos indivíduos em si, mas nas interações existentes entre eles, valorizando, assim, a perspectiva sistêmica<sup>31</sup>. Auroux (1998, p. 112) condensa as diferentes hipóteses discutidas até então da seguinte forma:

Qualquer que seja a gramática que você construa para explicar as produções de um grupo de sujeitos, admitindo que você possa verificar o maior grau de adequação desejado entre os *outputs* da sua gramática e os fenômenos, então se produzirá em dado momento fenômenos para os quais ela será inadequada, sem que com isso seja impossível de encontrar uma outra gramática adequada ao mesmo grau concernente a esses fenômenos<sup>32</sup>.

Isso não impede, no entanto, o uso do cálculo nas representações linguísticas, o que nos leva a duas conclusões importantes: primeira, o fato de se negar uma visão que coloque a criatividade no âmbito da recursividade linguística linear, como um sistema formal inato, não impede que os seres humanos sejam capazes de internalizarem raciocínios lógico-matemáticos para resolverem determinados problemas, como no caso da própria língua(gem). Sendo os algoritmos instrumentos técnicos (hipótese das ferramentas linguísticas), eles podem abandonar, em parte, a sua “natureza mental”, na medida em que seriam artefatos humanos criados e

---

<sup>31</sup> Sobre esse aspecto na Psicologia Cognitiva, vide Lévy (1993). Há também um potencial de aproximação dessa hipótese de Auroux (1998) com aquela desenvolvida por Tomasello (2003) a respeito da “sociogênese da inteligência humana” e da “catraca cultural”. Sobre o primeiro aspecto, “[c]ompreender as outras pessoas como agentes intencionais iguais a si mesmo torna possíveis (a) processos de sociogênese por meio dos quais vários indivíduos colaboram entre si para criar artefatos e práticas culturais com histórias acumuladas, e (b) processos de aprendizagem cultural e internalização por meio dos quais os indivíduos em desenvolvimento aprendem a usar e depois internalizam aspectos dos produtos criados pela colaboração entre co-específicos” (TOMASELLO, 2003, p. 19-20). Sobre o segundo, “[a] metáfora da catraca nesse contexto pretende dar conta do fato de que a aprendizagem por imitação (com ou sem instrução ativa) propicia o tipo de transmissão fiel e necessária para manter a nova variante dentro do grupo, proporcionando assim uma plataforma para as futuras inovações – com as próprias inovações variando em função de elas serem individuais ou sociais/cooperativas” (TOMASELLO, 2003, p. 54). Para uma análise comparativa entre os pressupostos da AD e os de Tomasello (2003) e Vigotski (2008), vide Morais (2015a).

<sup>32</sup> Citação em francês: “Quelle que soit la grammaire que vous construisez pour expliquer les productions d’un groupe de sujets, en admettant que vous puissiez vérifier le plus grand degré d’adéquation voulu entre les *outputs* de votre grammaire et les phénomènes, alors il se produira à un moment donné des phénomènes pour lesquels elle sera inadéquate, sans que pour autant il soit impossible de trouver une autre grammaire adéquate au même degré concernant ces phénomènes” (AUROUX, 1998, p. 112-3. Tradução nossa).

institucionalizados pelos diversos grupos do saber. Segunda, o fato de os modelos lógico-matemáticos não poderem dar conta da totalidade do funcionamento das línguas naturais, avaliadas sob a perspectiva de regras, axiomas e algoritmos finitos que tornam possível a infinidade de produções, não implica que a Linguística deva abandonar as tentativas de formalização simbólica por meio de modelos automatizados. Diferentemente disso, esse posicionamento significa apenas que esses modelos formais não podem ser reduzidos à forma da própria linguagem (AUROUX, 1998).

Como destaca Auroux (1998), a abstração lógica é um avanço para o campo das Ciências da Linguagem. Todavia, considerar a não redução dos processos formais ao funcionamento da linguagem nos permite abandonar uma série de “raciocínios metafísicos”, os quais mais atrapalham do que ajudam. Como vimos anteriormente, ao contrário do que postulam os racionalistas, as atividades de cálculo nas línguas naturais desempenham o oposto da criatividade humana. Elas não são, portanto, as responsáveis pela inovação das “formas”, mas, sim, pela manutenção delas, por aquilo que se fixa, se estabiliza e se convencia.

Em vez de postular a existência de regras homogêneas presentes em todos os falantes de uma dada língua, seria mais produtivo enveredar pela construção de modelos interativos<sup>33</sup> que possam se deter sobre as diversas

---

<sup>33</sup> A Linguística Cognitiva (LC) tem avançado no tratamento da linguagem como um sistema dinâmico, assemelhando-se à definição de hiperlíngua em Auroux (2010, p. 243-4. Itálicos do autor): “Os acontecimentos na hiperlíngua (aquilo a que chamaremos de ‘discursos’) mudam igualmente (mais ou menos) a estrutura. [...] Em qualquer situação, ela é esta realidade última que engloba e situa toda realização linguística e limita concretamente toda inovação. [...] A hiperlíngua é um sistema dinâmico, a subdeterminação das atividades linguísticas pelas gramáticas pode ser derivada de um resultado matemático bastante conhecido. Não há em geral projeção ‘canônica’ de uma dinâmica sobre um subespaço; em outros termos, não há aplicação de um espaço multidimensional sobre dois subespaços suplementares, tal que o dado das duas dinâmicas obtidas possa permitir reconstruir a dinâmica inicial. Há então forçosamente perda de informação quando se estuda um fenômeno desse tipo sobre subespaços de dimensão inferior, mesmo se se ajustam os resultados obtidos sobre cada subespaço”. No âmbito da LC, Cameron & Deignam (2009, p. 148) destacam: “O discurso é tido como resultante da interação de múltiplos sistemas complexos dinâmicos que operam em vários níveis e escalas cronológicas [...]. Os sistemas complexos incluem mentes/cérebros, linguagem e recursos conceptuais (CAMERON, 2003). Sistemas dinâmicos complexos são sistemas de elementos conectados e interagentes que se encontram em constante fluxo. Em tais sistemas não são apenas os elementos que se modificam com o tempo, mas também as relações entre eles e é isso que define um sistema como ‘não-linear’. Como resultado da dinâmica não-linear, algumas modificações no sistema se dão de maneira abrupta e drástica, à medida que o sistema passa pela assim chamada ‘mudança de fase’ para um padrão de atividade diferente (esses padrões de atividade às vezes são denominados ‘atratores’). Outras vezes, o sistema se modifica continuamente, sem mudança de fase. Ao passar por mudanças de fase, o sistema parece se ‘auto-organizar’, com novos padrões ou atratores se desenvolvendo, se estabilizando por algum tempo e, em seguida, se modificando novamente. Abordagens emergentistas dão destaque à mudança não-linear e ao comportamento auto-organizacional

competências dos sujeitos, haja vista que eles possuem histórias de vida distintas. Cada sujeito está submetido a uma espécie de confrontação com o tempo, de modo que está apto a desenvolver novas competências, a “aprender” novas regras e a se utilizar de novas estruturas linguísticas. Cada ato do discurso só pode ser compreendido, em última instância, a partir do seu enquadramento em uma sequência linguística historicamente determinada (AUROUX, 1998).

Além disso, consoante Auroux (1998), toda análise racionalista terminará por ceder a uma abordagem empirista, uma vez que deverá considerar a dinâmica histórica e as interações sociais, responsáveis por possibilitar a construção das línguas reais. No dia a dia, um número indefinido de emissões linguísticas – não descritível por uma abordagem estritamente extensional – é gerado pela interação de um conjunto de sujeitos “aptos” a se comunicarem com um grau considerado – jamais absoluto – de intercompreensão.

Para finalizar, se a cognição não se reduz ao indivíduo, podemos afirmar, então, que a própria criatividade não o faz, levando-nos a uma releitura radical desses conceitos. A instrumentalização do mundo pelos humanos torna-se possível pela própria relação que os sujeitos estabelecem com ele. Seria impossível contar, por exemplo, se não tivéssemos coisas a serem contadas, demonstrando que a relação entre os sujeitos e o ambiente permite a emergência, a difusão e a interiorização do próprio conhecimento. Não haveria, enfim, conhecimento sem mundo a ser conhecido, categorizado e organizado.

### **Considerações finais**

No presente texto, apresentamos os principais pressupostos teóricos do Empirismo Cognitivo, em especial, no que concerne à relação entre

---

dos sistemas complexos e focalizam a ‘emergência’ de novos padrões de atividade do sistema após uma mudança de fase. O processo de emergência através da interação não-linear dos elementos do sistema é muito diferente de uma modificação que pode ser explicada com base na interação de elementos tidos como fixos em relação uns aos outros”. Para uma tentativa de aproximação entre LC e AD, consultar Morais (2015b; 2017).

pensamento, linguagem e história. Nesse intuito, vimos que a linguagem se constitui na interface entre o Real, o Simbólico e o Imaginário (Sistema RSI), avaliamos o processo de compartilhamento das estruturas internalizadas pelo pensamento humano e expusemos as hipóteses das ferramentas linguísticas, da subdeterminação gramatical, da determinação histórica dos sentidos e da conjectura sociológica, tentando demonstrar, sempre que possível, a proximidade epistemológica destas com outros campos de estudos.

Cumpramos destacar que a presente análise dialoga diretamente com a nossa tentativa de desenvolver nos últimos anos alguns diálogos teóricos que possam interessar à construção de uma abordagem discursiva cognitiva e histórica, ancorada em um viés de determinação última do simbólico humano pelo imaginário (histórico), e não pelo biológico, que, se é fundamental ao ser humano como espécie, encontra-se em um estágio anterior de complexidade em relação às línguas naturais e à realidade social. Desse modo, mais do que o interesse no entendimento do biológico na formação histórica do ser, interessa-nos, no sentido inverso, a compreensão da plasticidade desse organismo em decorrência da sua organização cultural.

## Referências

- ALTHUSSER, L. Sustentação da tese de Amiens. In: ALTHUSSER, L. **Posições I**. Rio de Janeiro: Edições Graal, 1978.
- AUROUX, S. **A revolução tecnológica da gramatização**. Campinas, SP: Editora da Unicamp, 1992.
- AUROUX, S. **La logique des idées**. Collection Analytiques 6. Montreal et Paris: Bellarmin et Vris, 1993.
- \_\_\_\_\_. **La raison, le langage et les normes**. Paris: PUF, 1998.
- \_\_\_\_\_. **Filosofia da linguagem**. São Paulo: Parábola, 2009.

- \_\_\_\_\_. A hiperlíngua e a externalidade da referência. In: ORLANDI, E. P (Org.). **Gestos de leitura**: da história no discurso. Campinas, SP: Editora da Unicamp, 2010.
- BAKHTIN, M. M.; VOLOSHINOV, M. **Marxismo e filosofia da linguagem**: problemas fundamentais do método sociológico na ciência da linguagem. São Paulo: Hucitec, 1990.
- CAMERON, L; DEIGNAM, C. A emergência da metáfora no discurso. **Cadernos de Tradução**. Porto Alegre, nº 25, p. 143-167, jul/dez, 2009.
- CHOMSKY, N. **Que tipo de criaturas somos nós?** Petrópolis, RJ: Vozes, 2018.
- ELIA, L. **O conceito de sujeito**. Rio de Janeiro: Jorge Zahar Ed., 2007.
- FOUCAULT, M. **A ordem do discurso**. São Paulo: Loyola, 1996.
- GUILHAMOU, J. **La langue politique et la Révolution française**. Paris: Méridiens-Klincksieck, 1989.
- HENRY, P. **A Ferramenta imperfeita**: língua, sujeito e discurso. São Paulo: Editora da Unicamp, 1992.
- KRIPKE, S. A. **Naming and necessity**. Cambridge, Mass: Harvard University Press, 1980.
- \_\_\_\_\_. **Wittgenstein on rules and private languages**. Cambridge: Harvard University Press, 1982.
- LÉVY, P. **As tecnologias da inteligência**: o futuro do pensamento na era da informática. Tradução Carlos Irineu da Costa. Rio de Janeiro: Editora 34, 1993.
- MORAIS, A. R. A. Superar o sujeito cartesiano, recuperar a cognição: o diálogo possível entre a análise do discurso francesa e o externalismo cognitivo. **Anais do Seminário de Estudos sobre Discurso e Argumentação**, 2014. v. 1.
- \_\_\_\_\_. Pensamento, Sujeito e Alteridade: Uma Análise Discursiva Da Tradição Soviética Dos Estudos Da Linguagem. **Gláuks**, v. 15 n. 1, 2015a. Disponível em: <http://docplayer.com.br/130250045-Glauks-glauks-vicosa-v-15-n-jan-jun-2015.html>. Acesso em: 21/08/2020.

\_\_\_\_\_. A metáfora como elo entre a Análise do Discurso, o Realismo Experiencial e os Sistemas Complexos. **Anais V Congresso Internacional sobre Metáfora na Linguagem e no Pensamento**. Belo Horizonte: Fale/UFMG, 2015b. p. 254-269. Disponível em: <http://www.letras.ufmg.br/congressometafora/Morais.pdf>. Acesso em: 21/08/2020.

\_\_\_\_\_. O Problema Mente/Corpo na Análise do Discurso Francesa. In: Souza, M. J. A.; LIMA FILHO, M. M. (Orgs.). **Escritos de Filosofia II: linguagem e cognição**. 1ª ed. Maceió: Edufal, 2017, v. 2, p. 293-315.

\_\_\_\_\_. Da natureza humana: as perspectivas epistemológicas de Piaget, Chomsky e Foucault. In: Souza, M. J. A.; Lima Filho, M. M. (Orgs.). **Escritos de Filosofia III: linguagem e cognição**. 1ª ed. Porto Alegre : Ed. Fi, 2019, v. 3, p. 98-115.

PÊCHEUX, M. **O discurso: estrutura ou acontecimento**. Campinas, SP: Pontes, 2008.

PIATELLI-PALMARINI, M (Org). **Théories du langage, théories de l'apprentissage: le débat entre Jean Piaget et Noam Chomsky**. Paris: Éditions du Seuil, 1979.

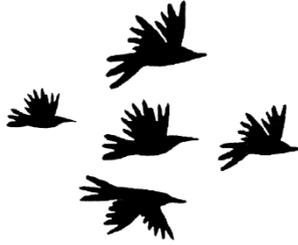
SAUSSURE, F. **Curso de Linguística Geral**. São Paulo: Cultrix, 2006.

TOMASELLO, M. **Origens culturais da aquisição do conhecimento humano**. São Paulo: Martins Fontes, 2003.

VIGOTSKI, L.S. **Pensamento e linguagem**. São Paulo: Martins Fontes, 2008.

WITTGENSTEIN, L. **Investigações filosóficas**. São Paulo: Editora Nova Cultural, 1999. (Coleção Os Pensadores)

ŽIŽEK, S. **Como ler Lacan**. Rio de Janeiro: Zahar, 2010.



## **Aproximações entre direito e moral nos conflitos entre normas**

*Juliele Maria Sievers*<sup>1</sup>

O presente capítulo pretende discutir a dicotomia entre as abordagens do direito positivo e do direito natural quanto às relações entre o Direito e a Moral. Nesse contexto, analisaremos como são avaliados os conflitos normativos de tipo “casos difíceis” segundo estas duas perspectivas: de um lado, a perspectiva positivista, que defende a autonomia do direito para a resolução dos conflitos e, de outro, o direito natural que apela a princípios morais para lidar com as situações de antinomia jurídica. Após verificar as limitações de ambas abordagens, analisaremos a alternativa apresentada através da chamada “justiça poética” de Martha Nussbaum (1997), defendendo que a noção de *empatia* seria determinante para a resolução de conflitos normativos e tomadas de decisão.

As aproximações entre os campos da Filosofia e do Direito são um campo de investigação já bem desenvolvido, cuja História remonta aos estudos sobre o direito Romano antigo, até questões atuais como aquelas envolvendo a legislação sobre aborto e eutanásia, por exemplo. A partir da

---

<sup>1</sup> Doutora em Filosofia pela Université Lille 3 – Charles de Gaulle. Professora adjunta do Curso e do PPG de Filosofia da Universidade Federal de Alagoas (UFAL). Contato: [juliele.sievers@ichca.ufal.br](mailto:juliele.sievers@ichca.ufal.br)

abordagem que se situa em um nível mais teórico (“o que é uma norma?”, “por que seguimos regras?”, “o que legitima uma ordem ou comando como sendo uma norma obrigatória?” etc.) desenvolveu-se inclusive a disciplina híbrida da Filosofia do Direito, que conta com um desenvolvimento histórico e teórico autônomo e bem delineado.

Uma das questões centrais da Filosofia do Direito trata da relação entre dois sistemas normativos que perpassam nossas atitudes e tomadas de decisão cotidianas, a saber, a relação entre o direito mesmo e a moralidade. As investigações teóricas quanto a esta relação resultaram em uma dicotomia entre, de um lado, a corrente que separa os dois sistemas entre si, defendendo a autonomia de cada um, que é a perspectiva do positivismo jurídico e, de outro lado, a corrente que defende a interdependência entre os dois campos como sendo complementares, que é defendida pela perspectiva do direito natural.

Frente a este contexto, os possíveis conflitos normativos que possam surgir em cada sistema normativo (Direito ou Moral) receberão um tratamento diferenciado, de acordo com as duas perspectivas mencionadas. É nesse sentido que avaliaremos um tipo especial de conflito, as antinomias do direito chamadas de *hard cases* ou casos difíceis, que são conflitos de deveres parecidos com os dilemas morais dentro do campo da Ética<sup>2</sup>. Nossa tese é a de que, enquanto no direito positivo existem métodos específicos para solucionar os casos difíceis, eles raramente são solucionados, na prática jurídica, sem o apelo à elementos externos, especificamente a princípios morais. Ao mesmo tempo, se o positivismo recai, uma vez aplicado na prática, numa perspectiva que se assemelha à de sua corrente “rival”, veremos que o naturalismo, ao defender suas teses, recai, por sua vez, muitas vezes, na defesa de princípios morais conservadores e inaceitáveis dentro de uma sociedade

---

<sup>2</sup> Quanto ao uso indiscriminado que fazemos aqui entre os termos “ética” e “moral”, seguimos a perspectiva de Murcho (2004) “[...] a pretensa distinção entre a ética e a moral é intrinsecamente confusa e não tem qualquer utilidade. A pretensa distinção seria a seguinte: a ética seria uma reflexão filosófica sobre a moral. A moral seria os costumes, os hábitos, os comportamentos dos seres humanos, as regras de comportamento adotadas pelas comunidades. Antes de vermos por que razão esta distinção resulta de confusão, perguntemo-nos: que ganhamos com ela?”.

cultural, econômica e socialmente múltipla. Ao final, veremos como a noção de *empatia* defendida por Martha Nussbaum (1997) oferece uma alternativa mais moderada para a solução dos conflitos normativos pelos juízes.

### **Os conflitos do direito: positivismo e naturalismo jurídicos**

A particularidade da tradição do direito natural é a concepção de que valores morais (bom, justo, louvável...) são não apenas qualificações, mas o próprio critério de definição do que seja uma norma legítima ou válida. Visto isso, cabe lembrar, desde já, que a “validade” de uma norma moral ou jurídica não tem nenhuma correlação com o termo “validade” tal como é usado no âmbito da lógica, direcionado para a correção de um argumento. No âmbito normativo do Direito, a norma é válida quando “vale”, ou seja, quando é obrigatória a sua observância pelos seus destinatários. Dito isto, temos que, no contexto do naturalismo jurídico, a norma “vale” se estiver de acordo com princípios de justiça, ou seja, se for uma “boa” norma, uma norma justa. Cumprido este critério, a norma “vale”, ou seja, seu cumprimento e aplicação são obrigatórios.

Podemos agora observar historicamente como o desenvolvimento da tradição do direito natural apresentou tais noções específicas – o justo, o bom – como sendo o critério de validade de suas normas, enquanto que, no âmbito do direito positivo, o único critério de validade era precisamente a própria legalidade, ou seja, o fato de a norma ter sido criada ou aplicada por uma autoridade jurídica e seguindo os ditames impostos pelo próprio direito vigente e, portanto, sem correlações com valores morais externos à própria ordem jurídica.

Assim, se formos traçar um breve recorte histórico, vemos como, na Antiguidade, pelas palavras de Aristóteles, lei natural e lei positiva coexistiam, sendo direcionadas para diferentes esferas da vida social: a primeira numa esfera pública e a segunda para o tratamento de questões privadas, mais específicas. De acordo com Aristóteles:

A lei universal é a lei da natureza. Pois, existe uma justiça natural e uma injustiça que é comum a todos, mesmo para aqueles que não têm associação ou contrato um com o outro. Como por exemplo, o mostra a Antígona de Sófocles ao declarar que é justo enterrar Polínicês, embora seja proibido: ela quer dizer que era justo por natureza. Não é de hoje, nem de ontem que essa lei existe, mas eterna: e ninguém sabe desde quando surgiu (ARISTÓTELES, *Ret.* 1373 b 12-13).

A passagem acima trata, implicitamente, da tensão entre a lei “comum”, tradicional que, nos termos do naturalismo jurídico, se trata como “eterna e imutável”, e a lei fruto do poder “transitório” do governante, ou seja, a lei que é fruto da sua vontade soberana e, portanto, arbitrária. A arbitrariedade e contingência da lei fruto da criação humana é característica da abordagem positivista do Direito. Enquanto o conflito é menos numeroso – na verdade, idealmente inexistente – entre normas de um mesmo sistema jurídico, ele pode ser identificado com mais facilidade quando colocamos uma norma tida como fruto do direito ancestral ou divino frente uma norma gerada por uma autoridade política.

Dito isto, vejamos então como esta relação se desdobra em outro momento no contexto da Antiguidade, tomando como exemplo, o caso dilemático apresentado na Antígona de Sófocles, já mencionado na passagem acima, de Aristóteles. Na história, Antígona, tendo seus dois irmãos mortos em combate, depara-se com a diferença de tratamento dado pelo representante do Estado, Creonte, rei de Tebas, em relação aos serviços fúnebres: enquanto Creonte destina a Etéocles todas as honrarias fúnebres por haver defendido Tebas, ele ordena que o corpo do irmão, Polínicês, seja deixado ao relento, como alimento aos pássaros, por haver atacado Tebas e traído o povo. Antígona desespera-se frente ao ato, já que os corpos não enterrados devidamente seriam destinados, segundo a tradição, a vagar eternamente numa espécie de limbo. Antígona então honra as chamadas “leis anteriores”, leis soberanas dos Deuses, ou seja, do direito “natural”, que garantem os serviços fúnebres a todos os seres humanos, e quebra a lei positiva promulgada pela autoridade de Creonte,

sendo então julgada por seu ato de enterrar o irmão. Antígona defende que não é “certo” ou “justo” negar ao irmão o direito universal ao sepultamento, fazendo com isso prevalecer a lei natural e pagando o devido preço ao ser condenada pelo tribunal por haver desobedecido a lei positiva do governante. Ela é condenada a ser sepultada viva em uma caverna, e Creonte se questiona, após a sucessão de infortúnios, sobre os limites de seu poder autoritário frente à tradição. Na Antígona, a lei do soberano não é seguida pela personagem porque:

*Não foi o meu Zeus que a proclamou / E nem a justiça dos deuses lá debaixo /  
Que fixaram aos homens as perenes leis. / Não penseis que teus decretos  
fossem tão fortes / A ponto que um mortal pudesse transgredir / As inescritas  
e indeléveis leis divinas. / Elas não são de hoje, nem de ontem, são eternas. /  
E ninguém nunca soube de onde elas vieram (SÓFOCLES, 2006, p. 450).*

Posteriormente, no período medieval, a norma “válida” é aquela reconhecida como proveniente da Razão divina, como fruto da vontade de Deus. Ela ainda é considerada como um valor eterno e imutável, estando, neste período, claramente em um patamar acima daquele da lei positiva. A lei “humana”, em oposição à lei “divina”, é relegada novamente apenas a aspectos particulares da vida, como a realização de contratos e transações econômicas. O ponto é que, neste período, não se cogita o conflito entre normas advindas destas duas esferas, pois o direito divino é indubitavelmente soberano.

A partir dos séculos XVII e XVIII, o direito “eterno e imutável” passa a ser encarado como aquele que está conforme os ditames da Razão, enquanto que o direito positivo passa a ser correlacionado com o desenvolvimento do Estado, sendo uma emanção do poder deste. Neste período, portanto, como destaca Bobbio (1995, p. 21):

*Um outro critério parece surgir aqui, critério que não mais se refere à fonte, isto é, ao modo pelo qual um ou outro direito é posto, mas ao modo pelo qual os destinatários vêm à conhecer as normas: o direito natural é aquele de que obtemos conhecimento através da razão, de vez que esta deriva da natureza*

das coisas; o direito positivo é aquele que viemos a conhecer através de uma declaração de vontade do legislador. Glück apresenta como exemplo de direito positivo o usucapião, porque não deriva da natureza das coisas mas é determinado pelo legislador, e, como exemplo de direito natural, o princípio ‘pacta sunt servanda’ e o dever do comprador de pagar ao vendedor o preço avençado.

A partir da modernidade, portanto, voltamos à real possibilidade de conflito entre os âmbitos normativos – naturalista e positivista. Os avanços sociais, econômicos e políticos, e o desenvolvimento das relações entre indivíduos realocaram o direito positivo em uma esfera central, onde o que se entende por Direito ou por jurídico é o que é ditado pelo direito positivo por excelência. Neste contexto, o direito natural perde seus traços de controle jurídico e funde-se com o sistema moral, sendo praticamente indistinguível frente às normas de conduta morais, fornecendo os seus princípios mais basilares. No entanto, veremos que, atualmente, é precisamente nas situações em que surge o conflito drástico – os *hard cases* – quanto às normas do direito positivo, que surge também o apelo aos princípios naturalistas como norteadores das práticas de tomada de decisão.

### **Conflitos normativos na atualidade**

Os casos dilemáticos tratados pelo Direito apresentam geralmente grande visibilidade, tornando-se casos paradigmáticos também no estudo, pelo Direito, dos conflitos entre normas jurídicas e suas possibilidades de resolução.

Como dissemos anteriormente, parece ser precisamente nos casos dilemáticos, e não nas situações de aplicação “normal” da lei, que vemos apagam-se as distinções entre o campo jurídico e da moral. Ou seja, mesmo estando atualmente sob a égide do direito positivo, que condena toda e qualquer relação entre sistemas jurídicos e sistemas morais, por ser aquele objetivo e neutro enquanto este é subjetivo e valorativo, parece que,

surgindo o dilema ou o conflito, as fronteiras entre os dois campos tornam-se mais maleáveis ou sombreadas.

Antes de passarmos à análise específica de alguns casos “reais”, vejamos brevemente como também os chamados dilemas morais, que nada mais são do que os conflitos normativos pertencentes ao âmbito moral, se “comportam” frente à heteronomia normativa representada pelo poder jurídico e os princípios éticos. Tradicionalmente e em linhas muito gerais, podemos fornecer uma definição de dilema moral como sendo *“uma situação (...) na qual um agente moralmente deve e pode fazer uma ação, e moralmente deve e pode fazer outra ação”* (Gowans, 1994, p. 47). Tentemos, portanto, perceber que, mesmo “do outro lado”, no âmbito da moral, também existe a influência do campo normativo do direito positivo. Analisemos o clássico dilema moral apresentado à personagem Sofia no romance de William Styron (1979), “A escolha de Sofia”.

Sofia é colocada frente à possibilidade de escolher qual de seus dois filhos será enviado ao campo de concentração ao invés da câmara de gás (sendo que, não havendo escolha, ambas crianças seriam enviadas à câmara de gás). Ela responde que “não pode”, ou “não consegue” escolher, mas acaba seguindo um princípio moral utilitarista que faz com que escolha o menino mais velho ao invés da menina, mais frágil, pensando que esta teria, em todo caso, maior dificuldade de sobreviver. Sofia segue pelo resto da vida completamente atormentada pela decisão.

Ora, tal situação de conflito apenas surge por conta da autoridade jurídica do oficial nazista que interpela a personagem. Ou seja, Sofia estava forçada pela autoridade a escolher, estava vinculada normativamente à vontade e ao poder autoritário do oficial. A personagem, mesmo obedecendo e efetuando a escolha, segue com o eterno arrependimento ou resquício da culpa inevitável, característica da existência de um dilema moral real e genuíno, e da chamada aflição moral caracterizada por Bernard Williams (1965).

Já voltando ao ponto de vista do Direito, temos que o positivismo realiza a correlação da legitimidade da norma com sua validade

unicamente, e não com seu conteúdo, ou seja, com o comportamento que ela obriga ou proíbe. Afirma Bruno Celano, acerca da teoria do maior expoente do positivismo jurídico, o jurista Hans Kelsen, que:

De acordo com Kelsen, validade é uma força mandatória (*binding*). Uma norma possui força mandatória se e somente se deve ser obedecida. Dizer de uma norma que deve ser obedecida remete a dizer que se deve comportar conforme a norma prescreve. Então, por exemplo, afirmar que a norma “Crianças devem obedecer a seus pais” é válida remete a afirmar que as crianças devem fazer o que a norma lhes prescreve. O que a norma lhes prescreve é, no entanto, que obedçam seus pais. Afirmar que a norma “Crianças devem obedecer seus pais” é válida remete a afirmar, portanto, que as crianças devem obedecer seus pais. De maneira geral, aquele que afirma que a norma ‘Op’ é válida está ao mesmo tempo afirmando que Op, e vice-versa (o enunciado “A norma ‘Op’ é válida” implica – e é implicada por – o enunciado ‘Op’). Segundo a concepção de Kelsen sobre a validade, portanto, afirmar que determinada norma é válida remete a (é logicamente equivalente a) afirmar – i.e. reiterar, aceitar, endossar, assentir, a norma ela mesma” (CELANO, 2000, p. 178).

Assim, para Kelsen e para o positivismo jurídico, afirmar a validade de uma norma jurídica não é conferir à norma a *propriedade* de ser válida, assim como enunciados têm a *propriedade* de serem ou verdadeiros ou falsos. A validade, sendo a própria existência da norma enquanto legítima (*binding*), é justamente a única definição possível da norma, e explicar o que seja “norma” significa o mesmo que justificar sua validade, o que é, por sua vez, feito simplesmente ao indicar-se que a norma foi corretamente criada, ou seja, que está amparada pelo ato de vontade legítimo de uma autoridade jurídica, o que remete a dizer que ela é o fruto da vontade de uma autoridade jurídica ou pessoa autorizada pelo Direito.

Cabe mencionar que, além dos casos de conflito normativo, os dilemas no Direito também surgem em outras circunstâncias, como no caso da falta de amparo jurídico para questões que surgem na sociedade em constante transformação. Um exemplo claro seriam os casos polêmicos

acerca da legislação em torno do melhoramento genético de embriões, entre outros casos envolvendo novas tecnologias.

Deste modo, se esta prerrogativa kelseniana permite que validem-se normas como aquela destinada à Sofia, ou seja, com conteúdos cruéis e arbitrários do ponto de vista moral, ela também previne um bom número de conflitos internos ao Direito, dado que este deve prezar pela coerência interna, evitando que surja uma nova norma obrigando, por exemplo, o que outra norma, anterior, proibia. Porém, os dilemas morais, como o de Sofia, encontram-se num patamar de dificuldade de resolução diferente dos meros conflitos normativos, eles se assemelham mais ao que é chamado, do Direito, de “casos difíceis”.

### **Aspectos morais nos casos difíceis do direito**

Os casos jurídicos ordinários envolvendo conflitos entre normas podem ser resolvidos através da aplicação de princípios regulativos do Direito, a saber, aplicação da lei derogatória que obedece aos princípios *lex posterior derogat priori* ou *lex specialis derogat generalis*, entre outros.

Um conflito entre duas normas existe quando o que uma norma fixa como devido é incompatível com o que a outra norma estabelece como devido. Resolve-se o conflito pelo fato de que a validade de uma das duas normas se extingue por uma norma derogante (KELSEN, 1986, p. 136).

Porém, quando a complexidade do caso é maior, normalmente a resolução depende, em última instância, do poder discricionário do juiz, ou seja, da sua capacidade de decisão através da interpretação, capacidade esta que lhe é conferida através de sua autoridade jurídica.

Nos casos em que não parece haver uma norma previamente existente sob a qual subsume-se o caso, ou ainda quando, frente à existência de mais de uma norma a respeito de um caso, não parece haver preferência de uma frente à aplicação da outra, surgem os chamados casos difíceis no Direito:

Nos casos difíceis, quando não é possível entre duas ou mais interpretações fazer prevalecer uma lei ou uma decisão anterior, o intérprete tem de fazer escolhas, pendendo para o lado daquela que pareça mais justa e aceitável. A decisão é complexa e vai refletir não apenas as opiniões pessoais do juiz sobre a justiça, a moral, a ética e a equidade, como também irá harmonizar esses ideais quando competem entre si (ATIENZA, 2002, p. 74).

De pronto observa-se que, mesmo frente às tentativas de “proteger” o âmbito jurídico de traços moralistas, é precisamente nos casos problemáticos como os conflitos entre normas, que notamos a proximidade entre o campo normativo da moral – com seus princípios reguladores – e o campo normativo jurídico, com seus casos difíceis.

Assim, frente à supracitada passagem de Atienza, uma nova pergunta emerge: quem ou qual princípio irá “harmonizar” as opiniões pessoais do juiz quanto aos casos em questão? Qual o critério a ser aplicado? A nosso ver, o juiz se encontra, nessa situação precisa, do mesmo ponto de vista decisório que o sujeito face a quem apresenta-se um dilema moral, ou seja, frente a um terreno inóspito de escolhas, onde não existem direcionamentos a seguir nem fórmulas a aplicar.

Vejamos um exemplo paradigmático do direito positivo, o chamado caso “The Queen vs. Dudley and Stephens”, ocorrido na Inglaterra em 1884. Trata-se do caso verídico envolvendo atos de canibalismo que teriam sido efetivados em circunstância de necessidade. O estado de “necessidade”, no Direito, demonstra uma situação em que o sujeito encontra-se em uma situação tão dramática que exclui-se toda responsabilidade civil, ou seja, o sujeito age temendo por sua própria vida, numa situação que não foi gerada por seus próprios atos e nem poderia ser prevista por ele ou por mais ninguém a seu alcance.

O processo em questão foi movido pela corte britânica para julgar Dudley e Stephens, respectivamente comandante e imediato da embarcação Mignonette. Após o naufrágio da embarcação, apenas quatro membros da tripulação sobreviveram, dentre eles os dois réus do processo, passando vários dias em um bote salva-vidas, sem comida ou água. Sem

previsão de resgate e em estado de desespero completo depois de 19 dias, Stephens mata o camareiro Parker, de 17 anos – já em estado de coma –, para que os três tripulantes restantes pudessem alimentar-se de seu corpo e beber seu sangue. Dois dias depois, os três sobreviventes são resgatados por um navio que os avista.

O processo que, obedecendo a lei sobre homicídio da época, resultaria em uma sentença de pena de morte, transformou-se numa sentença de prisão de 6 meses. O principal elemento a favor da redução da pena foi a óbvia particularidade das circunstâncias envolvendo o caso de necessidade, assim como o apoio da mídia e a comoção do grande público.

Novamente vemos, portanto, um elemento advindo da tradição do direito natural – ou da *Common Law*, no caso trazido – adentrando-se nos limites do direito positivo. Este caso mostra como, em situação extremas, volta-se a observar as confluências entre direito positivo e elementos de moralidade, quando se evoca um princípio como o da “necessidade”, que apela a uma situação de exceção à regra.

Cabe mencionar que a corrente do direito natural ainda apresenta autores e abordagens contemporâneas, que se opõem à solução “mecanicista” do direito positivo aos casos reais. Quanto aos casos de conflito, e fazendo frente à aplicação de princípios positivistas formais como o da derrogação, surge a corrente atual do direito natural que prega a aplicação dos chamados “bens humanos básicos”. John Finnis (2007) é atualmente o defensor desta corrente, vendo nestes princípios práticos o motor de ação correta. Assim, um dos sete bens humanos básicos que o autor elenca, a *vida*, seria o “pré-requisito” para o agir moral enquanto um engajamento na busca de sua preservação nos meios da saúde, por exemplo. Frente a vários ataques, Finnis não apenas mantém sua classificação dos bens, como ainda incorpora elementos básicos bem menos “evidentes” como o matrimônio, por exemplo, que é por ele especificado como devendo ser exclusivamente formado por pessoas de diferentes sexos:

Na perspectiva finnissiana, trata-se de um equívoco, dada a realidade da associação humana que é o casamento, configurando descompasso com os bens humanos básicos percebidos pela razão prática, em prejuízo ao florescimento humano, por fragilizar a instituição do casamento, estimular trajetórias de vida incapazes da associação típica, necessária e valiosa do casamento enquanto comunidade familiar baseada na união biológica (heterossexual, permanente e procriativa), tudo sem esquecer as consequências para a educação de crianças e adolescentes (RIOS, 2018, p. 12).

Similarmente, questões polêmicas ou ainda casos difíceis relativos, por exemplo, ao direito ao aborto seriam abordados pela perspectiva de John Finnis através do seu princípio de *garantia de acesso ao bem básico da vida*, o que o faria então unir-se ao coro dos movimentos conservadores opostos às políticas de proteção dos direitos das mulheres. Tais desdobramentos tornam a abordagem do autor extremamente problemática, parecendo evocar problemas além daqueles que se propõe resolver. Deste modo, os bens básicos de Finnis, facilmente associados a simples valores individuais que reforçam padrões de comportamento excludentes e discriminatórios, atentam ao próprio valor teórico e científico de sua abordagem da normatividade em geral.

### **Emotivismo e decisões pautadas pela empatia**

Temos, frente ao panorama traçado, uma perspectiva clara de como os casos difíceis, tipos especiais de conflitos entre regras do direito positivo vigente, abrem espaço para a aplicação de termos e conceitos advindos da tradição do direito natural. Os princípios do direito natural invocados na *Common Law*, por exemplo, representam ideais de justiça que seriam tomados como gerais (eternos e imutáveis) ao ponto de sua própria legitimidade ser de reconhecimento geral.

Ora, por mais intrínsecos e abstratos que tais conceitos morais possam parecer, não parecem lograr espaço legítimo dentro do campo positivista, conhecido, com Kelsen, pela “pureza” da teoria que propõe:

Quando a si própria se designa como “pura’ teoria do Direito, isto significa que ela se propõe garantir um conhecimento apenas dirigido ao Direito e excluir deste conhecimento tudo quanto não pertença ao seu objeto, tudo quanto não se possa, rigorosamente, determinar como Direito. [...] Porém, um relance de olhos sobre a ciência jurídica tradicional [...] mostra claramente quão longe ela está de satisfazer a exigência da pureza. De um modo inteiramente acrítico, a jurisprudência tem-se confundido com a psicologia e a sociologia, com a ética e a teoria política. Esta confusão pode porventura explicar-se pelo fato de estas ciências se referirem a objetos que indubitavelmente têm uma estreita conexão com o Direito. Quando a Teoria Pura empreende delimitar o conhecimento do Direito em face destas disciplinas, fá-lo não por ignorar ou, muito menos, por negar essa conexão, mas porque intenta evitar um sincretismo metodológico que obscurece a essência da ciência jurídica e dilui os limites que lhe são impostos pela natureza do seu objeto (KELSEN, 2003, p. 1-2).

De fato, o direito positivo parece operar de maneira “funcional” nas bases de seus preceitos teóricos cientificistas e objetivos. Porém, por que não é autônomo também nos casos problemáticos como o conflito de normas? Por que falha quanto ao alcance de seus princípios teóricos a respeito de questões práticas, como a tomada de decisão nos casos difíceis?

O Direito e a vida prática são permeados por situações de tomada de decisão, e é precisamente nestes momentos que revelamos as motivações por trás de nossas escolhas. Como afirma Aristóteles, na sua *Ética a Eudêmio* (1992, 1228 a), que:

[...] é a partir das escolhas deliberadas de alguém que julgamos o seu caráter, isto é, julgamos o caráter de alguém a partir daquilo em vista do que ele age e não a partir da sua ação [...]. E é observando as escolhas que os homens fazem, antes que seus atos, que nós os louvamos ou censuramos.

Talvez existam alternativas menos drásticas do que a recusa kelseniana dos elementos externos (princípios morais) para o sistema legal, ou ainda menos problemáticas do que a lista de elementos pré-morais de John Finnis, que acaba por nos guiar a um extremo conservadorismo. A teoria de John Rawls (1971), por exemplo, trazendo a

tese do equilíbrio reflexivo através de princípios universais históricos confrontados com os juízos ponderados de justiça, poderia ser uma via mais comedida a ser avaliada.

De todo modo, a “lição” que prevalece é a de que a existência dos – e a nossa dificuldade de lidar com – casos difíceis mostra a rigidez das duas abordagens frente a um sistema jurídico que está em constante mudança, pois se refere, por sua vez, a uma sociedade que está em constante mudança também. Se as únicas alternativas oferecidas aos casos até agora sugerem uma maior parcimônia na aplicação das teses, abrindo espaço para uma maior flexibilidade na tomada de decisão do agente, talvez outros princípios possam entrar de fato em vigor. No caso da canibalismo dos naufragos, a *empatia* em relação aos sobreviventes foi, ao que parece, um fator determinante.

Frente a isso, tomamos como princípio norteador a noção de *empatia* como abordada por Martha Nussbaum (1997) como um possível aliado do juiz na tomada de decisões:

O raciocínio técnico jurídico, o conhecimento da lei e as restrições da jurisprudência desempenham um papel central no bom julgamento, fornecendo os limites dentro dos quais a imaginação deve operar. (...) A concepção aristotélico-literária, combinada com as restrições institucionais, oferece um um complexo ideal de neutralidade jurídica que é uma rival poderosa frente outras concepções influentes desta norma (NUSSBAUM, 1997, p. 118).

Segundo a autora, a empatia seria uma faculdade promovida pela imaginação e cultivada através da aproximação com a literatura. A “sensibilização” daquele que toma as decisões seria um fator positivo a favor da melhor solução a ser tomada, de maneira, porém, neutra, desvinculada de valores individuais. Através de leitura de romances, por exemplo, e do exercício imaginativo acerca do que “poderia acontecer” em uma história fictícia, aprendemos lições sobre as possibilidades acerca de eventos importantes, assim como acerca da multiplicidade de perspectivas

a respeito de determinado fato exercitando nosso “olhar” a partir do ponto de vista de outra pessoa (personagens).

A imaginação literária é parte da racionalidade pública, mas não o todo. Creio que seria extremamente perigoso sugerir que o raciocínio moral regido por regras seja substituído pela imaginação empática. De forma alguma sugiro isso. Defendo a imaginação literária precisamente porque me parece um ingrediente essencial de uma postura ética que nos exige nos interessarmos ao bem estar de pessoas cujas vidas estão muito distantes da nossa. Esta postura ética deixa margem ao recurso a regras e procedimentos formais, incluídos os procedimentos inspirados pela economia. Pessoalmente, sustento uma versão da postura ética derivada de Aristóteles, porém tudo o que defendo aqui pode ser concebido dentro de um kantismo modificado para conferir às emoções um papel cognitivo claramente delimitado (NUSSBAUM, 1997, p. 18).

Assim, a perspectiva racionalista ou cientificista é substituída por uma abordagem pautada na imaginação e sua capacidade de veicular emoções cultivadas através do exercício da empatia. Sem elencar princípios determinados como Finnis, ou ainda uma moral específica (como parece fazer Finnis, privilegiando uma moral cristã conservadora), Nussbaum apresenta uma via interessante para o desenvolvimento de virtudes e habilidades de resolução de problemas como os conflitos.

De fato, seria preciso delimitar mais claramente o critério para o direcionamento da referida noção de empatia ou compaixão. Mas a posição da autora não vai até determinar precisamente, através das habilidades de se colocar no lugar do outro gerada pela leitura dos clássicos, qual atitude tomar. Mais complexa que isso, a ideia é que decisões envolvendo temas sensíveis podem ser melhor encaradas quando o agente possui uma maturidade afetiva gerada pelo cultivo, através da leitura de literatura, da empatia. Este exercício, mais do que qualquer outra coisa, lhe proporcionaria uma mais abrangente visão de mundo e dos outros indivíduos, de modo a guiar mais acertadamente sua decisão.

Assim, o romance, tal como o “Tempos Difíceis”, de Charles Dickens, por exemplo, seria, para Nussbaum, algo sobre o qual:

Se poderia dizer que seu tema consiste na interação entre as aspirações gerais humanas e certas formas particulares da vida social que estimulam ou frustram determinadas aspirações, modelando-as neste processo. O romance (ou pelo menos o romance realista do qual nos ocupamos) apresenta formas persistentes de necessidade e desejos humanos manifestados em situações sociais específicas. Estas situações normalmente diferem bastante das do leitor. O romance, reconhecendo isto, apela em geral a um leitor implícito que compartilha com os personagens certas esperanças, temores e preocupações gerais, podendo por isso formar laços de identificação e empatia com eles, mas que também vive em um âmbito distinto e visa informar-se acerca da situação concreta dos personagens. Desta maneira, a mesma estrutura de interação entre o texto e seu leitor implícito convida o leitor a perceber como os traços cambiantes da sociedade e suas circunstâncias afetam a realização – mais ainda, a estrutura mesma – das esperanças e desejos comuns a eles (NUSSBAUM, 1997, p. 32).

A proposta de Nussbaum seria de que o “ensino” do exercício da empatia seja efetuado já nas escolas de Direito, de modo a formar profissionais com habilidades, além de técnicas, humanistas e sensibilizadas. A respeito disso, concordamos, com a posição de Dunlop a respeito da “justiça poética” defendida por Nussbaum:

A questão de se a literatura ensina ou não a ética em algum senso representa um debate acalorado no criticismo literário do século XX. (...) Minha posição é de que a ficção muitas vezes levanta e examina, usualmente sem resolver, questões éticas e políticas sobre a ordem jurídica. A literatura pode apontar a complexidade de uma questão ética. Ela pode evocar uma variedade de soluções possíveis. O leitor pode explorar o problema através dos olhos de vários personagens vivendo através de uma situação concreta e, como resultado, sua sensibilidade e compreensão serão enriquecidas (DUNLOP, 1992, 71, nossa tradução).

Assim, concluímos que o fato de nos voltarmos à literatura e às obras como a própria Antígona, talvez seja um modo de nos desvencilharmos da rigidez “purista” do positivismo vigente, sem por isso recair em princípios morais igualmente rígidos como aqueles que o naturalismo jurídico

contemporâneo propõe. A educação moral pela leitura e o cultivo da empatia através deste exercício visa preparar um profissional mais apto a encarar situações como as descritas nos casos difíceis, sem envolver nelas crenças pessoais pautadas em sistemas morais determinados, mas sim exercendo uma visão de mundo abrangente, múltipla e compassiva.

## Referências

- ARISTÓTELES. **Eudemean ethics**. Books I, II and VIII. Oxford: Oxford University Press, 1992.
- ATIENZA, M. **As razões do direito**: teorias da argumentação jurídica. São Paulo, Landy, 2002.
- BOBBIO, N. **O positivismo jurídico**: lições de filosofia do direito. São Paulo: Ícone, 1995.
- CELANO, B. Kelsen's Concept of the Authority of Law. **Law and Philosophy**, Vol. 19, No. 2, Mar, p. 173-99, 2000.
- DUNLOP, C. R. B. Literature Study in Law Schools. **Cardozo Studies in Law & Literature**, 3, 1, p. 63-110, 1991.
- FINNIS, J. **Lei natural e direitos naturais**. São Leopoldo: Editora Unisinos, 2007
- FOOT, P. Moral Dilemmas Revisited. In: **Moral Dilemmas**. Oxford: Oxford University Press, 2002, p. 175-87.
- GOWANS, C. **Innocence lost**: an examination of inescapable moral wrongdoing. New York: Oxford University Press, 1994.
- HOJECKÁ, P. Regina v. Dudley and Stephens. **Common Law Review**, p.31-6, 2017.
- KELSEN, H. **General theory of norms**. Clarendon Press, Oxford, 2001 [1979].
- KELSEN, H. **Teoria pura do direito**. São Paulo: Martins Fontes, 2003 [1960].

MURCHO, D. **Ética e Moral**: uma distinção indistinta. Disponível em: <[http://criticanarede.com/fil\\_eticaemoral.html](http://criticanarede.com/fil_eticaemoral.html).> Acesso em: 29/10/2019.

NUSSBAUM, M. C. **Justicia Poética**. La imaginación literaria y la vida pública. Chile: Editorial Andrés Bello, 1997.

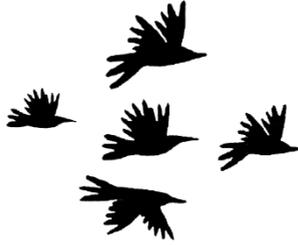
RAWLS, J. **A theory of justice**. Harvard University Press, 1971.

RIOS, R. R. Democracia, direitos humanos e direito natural: crítica à tese de John Finnis sobre o casamento homossexual. **Revista Publicum**, Rio de Janeiro, v. 4, n. 2, p. 11-36, 2018.

SÓFOCLES. **Antígona**. Trad. de Lawrence F. Pereira. Interpretação e notas de Kathrin H. Rosenfield. Rio de Janeiro: Topbooks, 2006.

STYRON, W. **Sophie's Choice**. London: Corgi, 1987 [1979].

WILLIAMS, B. Ethical Consistency. **Proceedings of Aristotelian Society**, 39, p. 103-24, 1965.



## **Explicações psicológicas em psiquiatria: o caso da neuropsiquiatria cognitiva**

*Rondineli Bezerra Mariano*<sup>1</sup>

*Richard Theisen Simanke*<sup>2</sup>

O lugar da teorização psicológica no domínio da psiquiatria constitui-se como um dos inúmeros problemas filosóficos com os quais essa disciplina historicamente se defrontou. Tal dilema tem sua origem na sua própria especificidade, quando concebida enquanto uma ciência médica dentro do esquema clássico do cartesianismo que atravessa a disciplina desde o começo do século XIX<sup>3</sup>. Ao tentar delimitar um objeto médico no terreno movediço do mental, a psiquiatria se vê diante de uma espécie encruzilhada ou, como bem sagazmente formulou o psiquiatra francês Henry Ey (1901-1977), um dos seus mais célebres intérpretes, se defronta com uma espécie de “dilema psiquiátrico”. Dualidade suicida, pois que ameaça dissolver a sua especificidade e identidade no quadro das disciplinas fronteiriças com as quais estabelece relação.

---

<sup>1</sup> Doutorando no programa de Pós-Graduação em Psicologia da Universidade Federal de Juiz de Fora (UFJF).

<sup>2</sup> Professor Titular do Curso de Psicologia da UFJF e docente da Pós-Graduação em Psicologia e Filosofia da mesma universidade.

<sup>3</sup> Henry Ey (2006a) situa o cartesianismo como a doutrina posterior a Descartes mais relacionado à Malebranche e outros filósofos do que propriamente ao filósofo francês.

Uma das alternativas para a disciplina, segundo Ey (2006b), é manter-se estritamente organicista e materialista, conformando-se ao modelo médico de doença e buscando a gênese dos estados psíquicos alterados numa alteração patológica do cérebro. No entanto, ao tomar tal partido dentro dos marcos cartesianismo, confunde-se com a patologia geral e acaba por relegar os estados psíquicos constituintes dos sintomas com os quais o clínico se depara ao status de epifenômeno, negligenciando a especificidade do psíquico e a experiência do sujeito. No entanto, ainda assim é necessário um mínimo de psicologia, pois os sintomas clínicos ainda se apresentam sob o plano psíquico, mesmo que a gênese seja concebida como totalmente orgânica (SIMANKE, 2002).

A outra opção é pensar a sintomatologia psiquiátrica simplesmente como alteração problemática da vida mental e crise psicovital desencadeadora de sofrimento. Mas, ao concentrar o foco na natureza propriamente mental do fenômeno, corre-se o risco de que a própria noção de “doença” perca sentido. A psiquiatria, assim, poderia perder seu status médico, dissolvendo-se num capítulo da psicologia geral que se ocupa das alterações da vida da mental que não possuem necessariamente uma natureza patológica. A natureza elusiva do objeto da psiquiatria fez com que uma incessante disputa entre teorias e terapias psicológicas e biológicas das afecções mentais se tornasse um tema constante no seu desenvolvimento histórico desde o final do século XVIII (PORTER & MIKALE, 1994).

No entanto, como o cartesianismo há muito tempo deixou de ser a única filosofia disponível para lidar com o problema mente/corpo, não é necessário que o debate sobre a relação entre explicações biológicas e psicológicas dos sintomas psiquiátricos se dê no interior dos seus marcos. Ainda mais quando o objetivo é obter alguma espécie de integração, no contexto psiquiátrico, entre teorias e análises psicológicas e biológicas. Várias abordagens em psiquiatria tentaram conciliar explicações biológicas e psicológicas e nosso objetivo é justamente apresentar uma perspectiva que se oriente nesse sentido. Assim sendo, apresentamos neste

capítulo de que maneira a neuropsiquiatria cognitiva<sup>4</sup> aborda os sintomas e as síndromes psiquiátricas e sua estratégia para integrar explicações psicológicas no escopo da psiquiatria por meio do desenvolvimento de modelos cognitivos de tais sintomas. Para essa disciplina, que se estabeleceu institucionalmente nos anos 1990 à imagem e semelhança da neuropsicologia cognitiva, a análise e teorização psicológica é uma etapa fundamental para a descoberta dos correlatos e redes neurais implicadas na manifestação sintomatológica dos quadros clínicos. Teríamos aqui um casamento razoavelmente feliz entre psicologia e neurociência, tão almejado pela psiquiatria?

Para entendermos a especificidade da neuropsiquiatria cognitiva, apresentamos inicialmente o ressurgimento da psiquiatria biológica, que se dá a partir dos anos 1970 (SHORTER, 2005) e sua reivindicação da exclusividade das explicações biológicas dos transtornos mentais. A partir daí, discutiremos como a NC toma forma opondo-se a essa perspectiva, ao defender a necessidade da construção de modelos cognitivos para explicação dos sintomas psiquiátricos. Por fim, apresentamos de que maneira é possível a integração entre explicações cognitivas e explicações formuladas na linguagem das neurociências por meio do recurso ao conceito de *mecanismo neurocognitivo*, corrente na neurociência cognitiva.

## A nova psiquiatria biológica

O retorno de uma orientação biológica como perspectiva dominante na psiquiatria americana e mundial inicia-se na década de 1970, desbancando as abordagens psicodinâmicas<sup>5</sup> que predominavam no *mainstream* psiquiátrico americano e ao redor do mundo desde o fim da Segunda Guerra Mundial. Para Shorter (1997), os principais

---

<sup>4</sup> Doravante NC.

<sup>5</sup> As teorias psicodinâmicas caracterizam-se pela ênfase posta na gênese psicológica do processo de adoecimento psíquico. Apesar da psicanálise ser uma das principais representantes dessas teorias, não é a única (ÁGUJAR, 2015).

acontecimentos responsáveis pela consolidação daquilo que foi então propagado como a ‘revolução biológica’ na psiquiatria foram 1) a descoberta de psicotrópicos eficazes, iniciada no início dos anos 50 (HEALY, 2002), para uma série de transtornos mentais e 2) a revelação do componente genético envolvido nessas condições, particularmente na esquizofrenia. A hipótese de que a causa dos transtornos mentais reside no desequilíbrio neuroquímico no cérebro surgiu justamente para explicar o sucesso dos fármacos em estabilizar os sintomas de uma série de transtornos depressivos, de ansiedade e psicóticos (WHITAKER, 2010). O velho adágio de que a doença mental não é nada mais do que uma doença do cérebro ganhava uma nova versão. Apesar de muito difundida, não há ainda marcadores biológicos nem ampla evidência que comprove de maneira inequívoca a existência de alterações neuroquímicas no cérebro das pessoas que apresentam sintomas psicopatológicos (HARRINGTON, 2019). Essa hipótese tornou-se popular e difundida justamente porque propunha uma explicação coerente para o relativo sucesso dos psicofármacos, onde antes não havia nenhuma, e por ser compatível com a perspectiva biológica dos transtornos mentais<sup>6</sup>.

A difusão dessas drogas foi decisiva para as pesquisas em neurociências, pois uma larga soma de investimentos foi ofertada pela indústria farmacêutica para investigar os efeitos das drogas no funcionamento do cérebro. O sucesso dessas drogas também estimulou o Instituto Nacional de Saúde Mental (NIMH) americano a financiar as pesquisas em neurociência e promover sua aliança com a psiquiatria (SHORTER, 2005). Uma poderosa aliada dessas pesquisas foram as tecnologias de neuroimagem que, segundo Walter (2013), foram de fundamental importância para a consolidação e o entusiasmo em torno dessa nova psiquiatria biológica. Até então, apesar de muitos psiquiatras acharem que subjacente aos transtornos mentais haveria um ‘*broken brain*’, não havia tecnologias para a visualização do funcionamento do

---

<sup>6</sup> Ver WHITAKER (2010), KIRSCH (2009) e CARLAT (2010) para uma apresentação crítica e contestação da hipótese de que a doença mental resulta de um desequilíbrio químico no cérebro.

cérebro desses pacientes. As novas tecnologias representavam agora a possibilidade de finalmente descobrir quais mudanças estruturais e funcionais no cérebro produziam os sintomas característicos dos transtornos mentais e como essa produção se dava (ANDREASEN, 2001). As primeiras dentre essas tecnologias, tal como a tomografia computadorizada, permitiam apenas a visualização das estruturas anatômicas do cérebro e motivaram a busca por mudanças estruturais, por exemplo, no cérebro dos esquizofrênicos. Já a segunda geração de técnicas de neuroimagem captava, além das estruturas, em alguma medida, o seu próprio funcionamento. A imagem por ressonância magnética funcional (fMRI) é uma das mais conhecidas entre elas, tornando-se muito utilizada nas pesquisas em psiquiatria a partir dos anos 1990.

A perspectiva biológica se tornou dominante no meio científico e profissional psiquiátrico, rivalizando com as perspectivas psicodinâmicas e psicossociais dos transtornos mentais ao denunciar a falta de rigor e de base científica para as explicações psicológicas da gênese dos sintomas nessas teorias, assim como a ineficácia de seus respectivos tratamentos (HARRINGTON, 2019). Não é de admirar, portanto, a exclusão da análise e teorização psicológica no seu escopo. Se os transtornos mentais resultam de algum tipo de patologia cerebral e há uma neurociência molecular, combinada com as tecnologias de neuroimagem, para observar e estudar o cérebro e assim elucidar as supostas alterações anatômicas, funcionais ou químicas, qual a razão para se deter em teorizações psicológicas abstratas? A ironia histórica é que, se o surgimento das tecnologias de neuroimagem levou a psiquiatria biológica a tomar distância dos modelos cognitivos das funções psicológicas, na neuropsicologia elas tiveram o papel exatamente oposto, tal como assinalado por Frith (2015). Essas tecnologias propiciaram o refinamento dos modelos cognitivos, por meio da busca de correlação entre padrões de lesão cerebral e déficits em domínios cognitivos específicos, tendo como arcabouço teórico os modelos

cognitivos que vinham sendo desenvolvidos pela psicologia cognitiva experimental desde a década de 1960.

O lançamento, em 1980, do DSM III (*Diagnostic and Statistical Manual of Mental Disorders - III*), manual de classificação psiquiátrica organizado pela APA (*American Association of Psychiatry*), foi outro fator decisivo para a consolidação da hegemonia da psiquiatria biológica. O DSM III rompeu com a influência que a psicanálise exercera sobre a elaboração do manual nas suas duas primeiras edições. Proclamava-se como ateuórico e alegava que as categorias clínicas estabelecidas tinham que ser validadas cientificamente. Os representantes da APA, desde então, celebram tal edição como um grande acontecimento e marco da reviravolta científica na psiquiatria (MAYER & HORWITZ, 2005). O otimismo no qual a psiquiatria estava envolta no início da década de 80 apostava que a descoberta dos correlatos neurais e dos genes relacionados às condições como esquizofrenia, transtorno bipolar e transtorno depressivo maior era apenas uma questão de tempo.

Em linhas gerais, os representantes dessa nova psiquiatria, tal como Andreasen (1984), advogavam que a disciplina estivesse fundamentada, tal como as outras disciplinas médicas, no materialismo biomédico (ZACHAR, 2000). O materialismo biomédico congrega um conjunto de visões sobre a psiquiatria que pode ser sintetizado através da tese de que o foco da disciplina deve ser o cérebro (GUZE, 1992). Como não existe nenhum evento mental normal ou patológico que não tenha como substrato o cérebro, o caminho mais seguro para encontrar terapias eficazes seria por meio da investigação do seu funcionamento. Assim, os transtornos mentais passam a ser abordados como doenças do cérebro, e a neurociência é eleita a principal aliada da psiquiatria, que se vê, dessa forma, mais bem integrada ao conjunto das ciências médicas.

Bem integrada e adotando os princípios científicos dessas outras disciplinas médicas, a psiquiatria finalmente conseguiria o mesmo sucesso e aperfeiçoamento nas suas intervenções e práticas que já tinham sido conquistados nesses outros domínios. É importante ter em conta, como já

assinalamos, que a psiquiatria, historicamente, sempre encontrou dificuldade em se posicionar no quadro das ciências médicas, justamente devido ao fato de que a apresentação sintomatológica dos seus quadros clínicos se manifesta como fenômenos mentais. Essa nova psiquiatria biológica, tal como outras perspectivas organicistas anteriores, resolve esse problema apelando para a fórmula de que, se os sintomas se manifestam no domínio do mental, o processo patológico que os desencadeia ocorre no cérebro, ou seja, trata-se de um processo biopatológico como aqueles responsáveis por qualquer outra doença, com a qual as disciplinas médicas se confrontem, tal como a diabetes, a leucemia ou o Alzheimer. Pois, por mais que as causas responsáveis pelos transtornos possam ser de natureza genética, familiar ou cultural, elas só se manifestarão alterando o funcionamento do cérebro.

Se essa psiquiatria apostava todas as suas esperanças nas neurociências como motor do seu progresso, não reservava lugar algum para as teorizações e abordagens psicológicas no quadro de uma psicopatologia científica. Quem elucidaria os transtornos psicológicos seriam as hipóteses da neurociência sobre o cérebro que, segundo Zachar (2000), são consideradas como hipóteses sobre algo que realmente existe, revelando o comprometimento desse materialismo biomédico com o realismo científico. Na verdade, a atitude reinante era de que o predomínio tanto das abordagens psicodinâmicas como do behaviorismo tinha representado um entrave para o desenvolvimento da psiquiatria como uma legítima ciência médica ao longo do século XX, como defendeu Andreasen (1984). Em suma, defendia-se que a análise psicológica dos transtornos mentais tinha atrasado o progresso científico da disciplina que, finalmente, estava sendo agora posto em marcha com o reconhecimento da sua base biológica e sua investigação nesses termos. No que concerne à perspectiva psicanalítica, realmente havia uma oposição explícita à concepção estritamente médica dos transtornos psicológicos. Se para a medicina é imprescindível desenvolver um sistema nosológico rigoroso, traçar fronteiras claras entre o patológico e o não

patológico e desvendar a base biológica dos transtornos, para a psicanálise nenhum desses objetivos é compatível com sua abordagem da vida psíquica.

Se, para os psiquiatras, todos os mistérios em torno das aflições mentais se solucionariam por meio da investigação do cérebro e suas disfunções, então o progresso científico em psiquiatria requeria deslocar o foco das teorias psicológicas para as biológicas. Quanto mais se conhecesse o intrincado funcionamento do cérebro, mais as hipóteses e teorias psicológicas se tornariam irrelevantes no âmbito da investigação científica em psiquiatria, relegadas, no máximo, ao status de metáforas ou ficções sobre processos biológicos reais (ANDREASEN, 2001). Assim, a identidade dos psiquiatras não mais se confundiria com a obscura figura do psicoterapeuta, mas antes se equipararia com a de médicos tratando verdadeiras disfunções orgânicas que, por contingência, manifestam-se no encontro clínico como eventos psíquicos. Esse tratamento seria conduzido por meio de psicofármacos ou outras terapias que agissem diretamente sobre o cérebro.

A atitude dessa psiquiatria para com os conceitos e teorias psicológicas revela certa semelhança com a posição do materialismo eliminativo em relação à psicologia. Ambos defendem que, com o desenvolvimento de uma neurociência madura, conceitos psicológicos tais como mecanismo de defesa, crenças nucleares, força do ego etc. se revelarão tão fictícios e desprovidos de poder explicativo para a compreensão das psicopatologias como a ideia de que a loucura era causada por algum tipo de possessão demoníaca, popular por tantos séculos. Em suma, nem a tese da possessão demoníaca, nem os conceitos psicológicos, na verdade, se referem a qualquer entidade ou processo real no mundo que possa explicar os sintomas psiquiátricos. A posição de Churchland (1995) – um dos principais defensores do materialismo eliminativo – em relação aos transtornos mentais aproxima-se bastante desses psiquiatras biológicos. Para eles, o sofrimento das pessoas que

experimentam algum tipo de transtorno mental só será realmente mitigado quando forem descobertas suas causas biológicas.

O que essa psiquiatria biológica reivindica, mesmo que não seja tão explícita nos argumentos quanto os materialistas eliminativos, é que a única forma legítima de explicação dos fenômenos psicopatológicos é através da linguagem das neurociências. Ou seja, o funcionamento da mente humana, incluindo seus estados alterados, vai ser desvendado não pela psicologia, mas por diferentes disciplinas biológicas. Na contramão de filósofos da psiquiatria como Murphy (2006), que defende estratégias explicativas pluralistas e integrativas para a disciplina, o programa dessa psiquiatria se compromete com a exclusividade de explicações neurofisiológicas dos fenômenos psicopatológicos, consideradas como as únicas legítimas. Gerrans (2014) denomina essa posição como a tese da autonomia explicativa. Toda a explicação das psicopatologias seria exaurida pela investigação dos mecanismos neurobiológicos envolvidos, não sendo necessário recorrer à configuração psicológica dos transtornos. Gerrans critica essa psiquiatria biológica absolutista em nome de uma abordagem que integre modelos e teorias cognitivas às pesquisas em neurociência, sem deixar de lado a investigação fenomenológica da experiência própria aos diversos sintomas psiquiátricos. A crítica à perspectiva biológica feita pela NC também se orienta nessa direção.

### **A NC e os modelos cognitivos dos fenômenos psiquiátricos**

A abordagem que a neuropsiquiatria cognitiva<sup>7</sup> propõe para os sintomas psiquiátricos é um contraponto ao otimismo ingênuo dos psiquiatras biológicos, alertando para a necessidade de compreender e teorizar esses fenômenos em termos psicológicos (DAVID, 1993). Mais especificamente, propõe que isso seja feito a partir dos modelos, métodos e técnicas de investigação da neuropsicologia cognitiva. Esta última

---

<sup>7</sup> Foi o psicólogo do País de Gales Hadyn Ellis (1945-2006) que fez o primeiro uso público do termo neuropsiquiatria cognitiva durante um simpósio no instituto de psiquiatria da Universidade de Oxford em 1991 (DAVID, 1993).

disciplina serviu de modelo para o desenvolvimento da NC que adota, em linhas gerais, os mesmos pressupostos. O que caracteriza a neuropsicologia cognitiva é a observação sistemática do funcionamento cognitivo intacto e alterado, por meio de testes cognitivos, tanto de pessoas que sofreram danos cerebrais anatômicos ou funcionais como de pessoas saudáveis, com o objetivo de compreender melhor os diversos sistemas de processamento de informação humano constituintes da cognição, assim como as interações entre eles (EYSENCK & KEANE, 2017).

Parte-se, de início, dos modelos cognitivos do funcionamento da memória, raciocínio, compreensão e produção da linguagem oral etc., para explicar os déficits e padrões de falhas característicos encontrados nesses domínios em alguns indivíduos. Por outro lado, os dados advindos dos testes cognitivos realizados com essas pessoas possibilitam, por sua vez, testar e aperfeiçoar esses mesmos modelos. Por exemplo, um tipo de falha específica na compreensão da fala, que não era prevista por um modelo cognitivo desse processo, possibilita a reformulação e correção do modelo para explicar tal ‘anomalia’. A hipótese é que, ao observarmos os déficits na compreensão e produção da fala característicos, por exemplo, das síndromes afásicas à luz dos modelos cognitivos, haverá um melhor entendimento do processamento da linguagem não alterado. Esses distúrbios são como experimentos naturais que possibilitam aos teóricos melhor compreenderem a arquitetura cognitiva humana. Para Coltheart (2008), a neuropsicologia cognitiva, apesar de fazer uso de estudos de neuroimagem, é um subcampo da psicologia cognitiva e não da neurociência, já que o intuito é utilizar a observação dos déficits cognitivos em módulos específicos para testar e pôr à prova os modelos cognitivos sobre o seu funcionamento.

Portanto, se os padrões de déficits decorrentes de alterações cerebrais estavam sendo bastante úteis para o desenvolvimento e correção de modelos cognitivos dos mais variados processos psicológicos, porque não tentar compreender a variedade dos sintomas psiquiátricos a partir desse mesmo *framework*? Os neuropsiquiatras cognitivos não estavam negando

em bloco o programa *mainstream* da psiquiatria biológica. Mas, como defende Frith (1992), apenas achar uma correlação entre alterações no sistema dopaminérgico do córtex pré-frontal e sintomas psicóticos como delírios e alucinações auditivas verbais não funciona como uma explicação satisfatória. Por que alterações nessa rota neuroquímica desencadeiam delírios e não outros fenômenos mentais patológicos? Para responder essa questão, é necessária uma explicação em termos cognitivos que dê conta das supostas falhas cognitivas implicadas no aparecimento dos delírios.

A análise e explicação em termos psicológicos do que acontece no interior dos sistemas de processamento de informação e na interação entre eles para que tais fenômenos venham à tona é, assim, um procedimento indispensável, e não uma metáfora a ser descartada por uma neurociência amadurecida. Pois como encontrar os correlatos neurais de sintomas psicopatológicos, se não se compreende em detalhes o funcionamento dos processos cognitivos que lhes são subjacentes, e tampouco sua estrutura psicológica? O argumento vai na mesma linha dos neuropsicólogos cognitivos que insistem que sem um modelo cognitivo preciso do funcionamento do sistema e subsistemas da memória, por exemplo, a tarefa de encontrar suas bases neurais está fadada ao fracasso. Isso equivaleria a procurar o substrato neural de algo que não se conhece suficientemente bem. Nesse sentido, a pesquisa neurocientífica dos processos cognitivos é dependente de uma teorização cognitiva rigorosa.

Dito de outra forma, a teorização cognitiva é indispensável, e a NC reabilita a análise do fenômeno mental em psiquiatria, ao mesmo tempo em que o aborda numa chave muito diferente das perspectivas psicanalíticas, que por tanto tempo foram hegemônicas no cenário psiquiátrico. Em linhas gerais, ela está alinhada ao movimento cognitivista em psicologia, que compreende a mente como um conjunto de sistema de processamento de informação conectados e organizados hierarquicamente (GARNHAM, 2009). O estilo e a linguagem de teorização sobre o mental da psicologia cognitiva receberam muita influência das ciências da computação, da teoria da informação, da

cibernética e de inovações tecnológicas tais como o computador digital (CASTAÑÓN, JUSTI & ARAÚJO, 2014). É desse conjunto de fatores que surge a metáfora computacional postulando a analogia entre mente e *software*, por um lado, e entre cérebro e *hardware*, por outro. Essa famosa metáfora está na base do funcionalismo, tal como elaborado originalmente por Putnam (1961) e, mais tarde, por Fodor (1968), enquanto tese fundante em filosofia da mente do movimento cognitivista em psicologia.

Tal metáfora permite compreender a relação entre mente e cérebro de uma nova maneira e legitima um nível de análise do mental autônomo e não necessariamente conectado e dependente da investigação de sua implementação física no cérebro. É verdade que a maioria desses teóricos tomam partido por uma posição materialista, assumindo que funções psicológicas são realizadas e implementadas pelo cérebro. No entanto, eles podem se restringir a um nível abstrato de análise dos processos cognitivos compreendidos em termos de fluxo de informação e transformações das representações mentais num dado módulo cognitivo. O teórico cognitivista se detém nas operações algorítmicas que os *inputs*, após serem codificados como representações, sofrem no interior de um dado sistema cognitivo, e nas suas sucessivas transformações que vão resultar num determinado *output*. Como aponta Thagard (2012), a estrutura e a composição do sistema cognitivo como um todo são abordadas da perspectiva do programador e não da sua implementação física.

A tarefa que a NC se atribui é justamente utilizar esses modelos cognitivos para explicar a gênese dos fenômenos psicopatológicos dos diversos transtornos mentais, tal como explicitado por dois dos principais formuladores da NC enquanto disciplina:

A neuropsiquiatria cognitiva começa com uma série de proposições e simplificações: vamos tomar os sintomas (a voz crítica desencarnada; os enunciados desarticulados; a falta de sentimento pelos outros; etc.) e os tratar como se fossem uma anormalidade neuropsicológica (uma afasia; uma agnosia; uma mudança de personalidade; etc.). Vamos tomar o corpo de

pesquisa biológica em psiquiatria que revela achados importantes sobre os transtornos e tratar como se fossem as lesões observáveis e bem demarcadas usadas nos estudos de caso em neuropsicologia cognitiva. Finalmente, vamos *usar os melhores modelos cognitivos para o conjunto das funções psicológicas normais considerando que os fenômenos psiquiátricos estão dentro do seu âmbito*. O resto fica a cargo do estudo empírico (DAVID & HALLIGAN, 1996. p. 2).

Em outro trabalho, os autores afirmam:

O ponto de início é um sintoma ou uma síndrome particular, seguido pela proposição de um sistema cognitivo plausível implicado em suas manifestações. Por último, inferências sobre a disfunção do sistema cognitivo correspondente são feitas baseadas em evidência clínica, teórica ou experimental (HALLIGAN & DAVID, 2001. p. 210).

Colocar a ênfase da investigação em um sintoma específico ou numa síndrome bem delimitada, e não em categorias diagnósticas amplas, tais como transtorno bipolar, esquizofrenia ou outra qualquer estabelecida pelos manuais oficiais de classificação psiquiátrica, é outro fator de distinção em relação à psiquiatria biológica. Se as categorias clínicas estabelecidas pelo DSM III em 1980 foram saudadas à época como um grande avanço científico para a psiquiatria, pois possibilitariam a descoberta de sua base biológica, no decorrer das décadas seguintes, instalou-se um crescente ceticismo quanto a sua validade científica e sua utilidade para a descoberta do processo patogênico que desencadeia os sintomas. Os transtornos mentais são muito heterogêneos na sua manifestação sintomática (BENTALL, 2005) e tomam formas distintas em diferentes pessoas. A diversidade é tanta que duas pessoas podem receber o mesmo diagnóstico de esquizofrenia ou depressão e ainda assim possuir poucos sintomas em comum (MURPHY, 2010)<sup>8</sup>.

---

<sup>8</sup> No caso da depressão, por exemplo, é provável que dois sujeitos percam o interesse nas atividades que outrora lhes davam prazer e apresentem o humor deprimido. Mas, enquanto um pode tornar-se agitado, sem apetite e com dificuldades para dormir, o outro pode comer e dormir além do habitual e ter ruminções obsessivas de culpa e pensamentos suicidas (MURPHY, 2010).

Desse modo, não faz muito sentido procurar pelo processo patogênico e desenvolver um modelo cognitivo que explique o transtorno bipolar ou a esquizofrenia como um todo. É mais viável focar nas falhas, déficits cognitivos e desregulação emocional relacionados, por exemplo, ao estado de mania, no caso do transtorno bipolar. Da mesma forma, não se procura por um modelo cognitivo geral da esquizofrenia. Antes, o foco da investigação está nos sintomas esquizofrênicos característicos, tais como os delírios persecutórios, a desorganização do pensamento e da fala, as alucinações auditivas verbais e a alteração do comportamento motor, entre outros. Os modelos cognitivos devem ser desenvolvidos para dar conta desses fenômenos específicos.

O risco em procurar pelos correlatos neurais da esquizofrenia ou de qualquer outro transtorno mental estabelecido de acordo com os critérios do DSM seria que:

A tão chamada “psiquiatria biológica” vem frequentemente tentando associar um diagnóstico clínico a déficits bioquímicos sem dar atenção ao nível de análise sistêmico que poderia fazer a mediação entre o comportamento e o substrato material... O perigo evidente é que essa abordagem levará a longas (e custosas) pesquisas pelo substrato biológico de entidades não existentes (HALLIGAN & MARSHALL, 1996, p. 5).

A crítica de Halligan & Marshall (1996) à psiquiatria biológica, além de apontar para a necessidade já mencionada de teorias sobre processos internos (cognitivos) para que se possa estabelecer uma ponte entre os sintomas psiquiátricos e os seus correlatos neurais, se endereça também ao problema do estatuto de realidade das categorias clínicas da psiquiatria. Para eles, essa psiquiatria, explícita ou implicitamente, considera os transtornos mentais como se fossem tipos naturais (*natural kinds*), que existem objetivamente no mundo e que são descobertas e classificadas de maneira objetiva pela psiquiatria<sup>9</sup>. Mas condições como esquizofrenia,

---

<sup>9</sup> Para uma discussão sobre a realidade dos transtornos mentais e o melhor modo de conceber seu estatuto ontológico, ver Zachar (2000; 2015) e Banzato & Zorzanelli (2014). Uma interessante alternativa para essa concepção realista das categorias diagnósticas, muito discutida pela filosofia da psiquiatria, é o conceito de *interactive kinds* (classes

transtorno delirante, transtorno do espectro autista (TEA) não são objetos naturais como tigres e árvores ou doenças como a tuberculose em que o processo etiopatogênico já está devidamente estabelecido, assim como as fronteiras de sua manifestação sintomática em relação a outras doenças e o seu curso evolutivo.

Decididamente, os transtornos mentais não são tipos naturais (Zachar, 2000). Com raríssimas exceções, não conhecemos com precisão os fatores de ordem biológica, social e psicológica que os desencadeiam – ou seja, sua etiopatogênese. Por isso, para um autor como Bentall (2013) que, na esteira dos neuropsiquiatras cognitivos, defende a incorporação de modelos e da teorização cognitiva em psiquiatria, a investigação da gênese desses sintomas não deve ser orientada por tais categorias clínicas amplas tradicionais<sup>10</sup>. Somente com a correlação entre déficits e vieses cognitivos, detectados por meio de testes e de avaliação psicológica subjacentes a sintomas específicos bem circunscritos, é que se pode descobrir as disfunções em um sistema de processamento de informação que produz tal sintomatologia.

É evidente que tentar explicar sintomas através de disfunções em módulos cognitivos não implica que, para cada sintoma psicopatológico específico, haja um e somente um módulo envolvido, e não uma interação entre eles. Ou ainda que um dado sistema de processamento de informação não possa estar relacionado a uma série de sintomas diferentes. Presume-se, por exemplo, que diferentes tipos de falhas e disfunções no sistema responsável por processos cognitivos de

---

interativas) introduzido inicialmente pelo filósofo Ian Hacking. Sinteticamente, uma classe interativa é “aquela que é criada ou significativamente modificada uma vez que um conceito tenha sido formulado e que se aja de acordo com ele de certa maneira. As classes interativas podem também retroagir (*loop back*) para influenciar nossos conceitos e classificações” (KHALIDI, 2010, p. 335). Outro aspecto importante das classes interativas é o efeito retroativo (*looping effect*) que podem ter sobre o comportamento dos sujeitos: um paciente pode passar a se comportar de forma diversa ao ser diagnosticado e tratado como esquizofrênico, seja para corresponder ou para escapar a essa classificação.

<sup>10</sup> Ironicamente, até mesmo os atuais representantes da psiquiatria biológica como Insel (2013) reconhecem que orientar as pesquisas a partir das categorias estabelecidas pelo DSM não tem sido produtivo para a disciplina, justamente porque essas categorias carecem de validade científica. Disso resulta um intenso mal-estar e desconfiança da profissão psiquiátrica com os sistemas de classificação oficiais dos transtornos mentais como o DSM 5 e o CID 10 (Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde) (HARRINGTON, 2019).

metarrepresentação conhecido como teoria da mente (ToM) expliquem tanto alguns sintomas do transtorno do espectro autista como aqueles característicos da esquizofrenia e de outros transtornos psicóticos (Harrington, Siegert & McClure, 2005). Entre as funções desse módulo está a atribuição de estados mentais tais como crenças e desejos a outras pessoas, a fim de explicar, prever, justificar, em suma, de tornar inteligíveis seus comportamentos. Esse módulo opera com metarrepresentações (representações das representações mentais dos outros e de si mesmo), possibilitando o desenvolvimento de uma teoria da mente baseada no pressuposto de que estados mentais têm eficácia causal sobre os comportamentos.

Falhas no funcionamento nesse módulo são associadas às dificuldades de interação e comunicação social presentes no espectro autista (BARON-COHEN, LESLIE & FRITH, 1986; BARON-COHEN, 1995). Justamente pela dificuldade em atribuir intenções aos comportamentos alheios, os pacientes dentro do espectro autista apresentam dificuldades em compreender situações sociais não rotineiras e complexas, tais como aquelas que envolvem fingimento, ironia, construções linguísticas metafóricas etc. Na esteira desses estudos, Frith (1992) propôs que ao menos parte da explicação de vários sintomas componentes do quadro da esquizofrenia pode ser atribuída a déficits e disfunções na teoria da mente<sup>11</sup>. Por exemplo, comprometimentos no planejamento da ação voluntária que levam à apatia e ao comportamento bizarro resultariam da falha do indivíduo em monitorar seus estados internos e de perceber onexo causal que eles estabelecem com seus próprios comportamentos. Os problemas de automonitoramento e déficits na apreensão dos próprios pensamentos que surgem espontaneamente na consciência, por sua vez, ajudariam a explicar em parte os delírios de controle externo e as alucinações auditivas etc. Enfim, para o autor, diferentes tipos de

---

<sup>11</sup> A abordagem de Frith dos sintomas que compõem o quadro clínico da esquizofrenia presente no seu livro *Cognitive Neuropsychology of Schizophrenia* é um ótimo exemplo do que se designa como neuropsiquiatria cognitiva.

*breakdown* no sistema de metarrepresentação compõem parte importante da explicação dos sintomas psicóticos.

Por razões de espaço, apresentamos a proposta de Frith resumidamente. Obviamente, um bom modelo cognitivo de um sintoma psiquiátrico, ou de um conjunto desses sintomas, precisa identificar a natureza específica da disfunção no sistema cognitivo em questão que resulta nos déficits expressos pelos sintomas, como também indicar se há participação ou não de outros domínios cognitivos na produção dos mesmos. Além disso, exige-se que a descrição do sistema seja feita de acordo com os requisitos de uma explicação mecanística (CRATSLEY & SAMUELS, 2013), em que todas as partes e subpartes componentes, bem como suas operações e interações, sejam especificadas em detalhes.

Compreendemos que a associação feita pela psiquiatria biológica nos anos 1980 entre análise e teorização psicológica dos transtornos mentais, por um lado, e falta de rigor e cientificidade, por outro, não se sustenta facilmente, para além de seu aspecto retórico e programático. A proposta da NC revela a viabilidade e a necessidade de abordar os sintomas psiquiátricos de um ponto de vista psicológico, nos marcos de uma abordagem experimental que não dispensa a análise clínica rigorosa, ao contrário, a exige para a elucidação dos meandros das manifestações psicopatológicas. Mas como se dá a integração entre a explicação cognitiva e a explicação na linguagem das neurociências?

### **Integrando modelos cognitivos e explicações neurais**

No esquema clássico da neuropsicologia cognitiva, o desenvolvimento de modelos cognitivos dos processos psicológicos não deve necessariamente ser sensível às informações e teorias sobre os processos neurais que os implementam. Tendo-se em vista os preceitos funcionalistas subjacentes à disciplina, os algoritmos relacionados ao processamento da visão ou da memória, por exemplo, poderiam ser testados por meio de simulação computacional e implementados, em

princípio, não apenas em redes neurais humanas, mas também num computador digital, numa rede neural artificial ou em qualquer outro meio físico adequado para servir de suporte ao comportamento inteligente. No entanto, a neurociência cognitiva critica essa abordagem puramente funcional e abstrata dos fenômenos mentais e reivindica que as descobertas sobre o funcionamento do cérebro pelas neurociências são valiosas e indispensáveis na compreensão das partes componentes da arquitetura cognitiva humana. Ou seja, é necessário construir teorias cognitivas à luz da compreensão dos mecanismos neurais concretos que implementam tais processos.

Essa postura é corrente entre cientistas, tais como Shacter (2008), que buscam orientar a teorização a partir tanto dos métodos da neuropsicologia cognitiva quanto dos dados disponíveis sobre o funcionamento de mecanismos cerebrais, atuando assim nas duas frentes de investigação. A percepção crescente, ao longo dos anos 1980, da existência de um *gap* entre a teorização cognitiva e as pesquisas em neurociência foi justamente o fator desencadeante do desenvolvimento da neurociência cognitiva como disciplina-ponte entre esses dois domínios de investigação (BECHTELL, 2009). Há, na verdade, uma ênfase cada vez mais reiterada para que a construção de teorias por parte tanto da psicologia cognitiva como da neuropsicologia seja orientada pelos achados das neurociências (EYSENK & KEANE, 2017; WRIGHT & BECHTEL, 2007).

Trazendo essa discussão para o escopo da NC, compreendemos que não é viável essa disciplina vindicar a autonomia explicativa cognitiva em relação ao nível de investigação e explicação neural, tal como fazem alguns dos mais eminentes representantes da neuropsicologia cognitiva, como Coltheart (2010), por exemplo. É mais interessante perscrutar uma via de acomodação entre teorias e modelos cognitivos e as pesquisas em neurociência que vão desde o nível intracelular até os níveis de funcionamento cerebral mais complexos e holísticos, caso se queira um arranjo razoável entre psicologia e biologia no interior da psiquiatria. É

nesse sentido que Murphy (2006) insiste na integração da psiquiatria com a neurociência cognitiva e Gerrans (2014) propõe que, em vez de concebermos a NC como uma neuropsicologia cognitiva dos sintomas psiquiátricos, deveríamos apostar numa espécie de “psiquiatria neurocognitiva”. Para além de uma simples disputa terminológica, há um conjunto de pressupostos subjacentes nessa redefinição. Alguns expoentes da NC exigem total autonomia na construção dos modelos cognitivos em relação ao nível neural, seguindo o modelo clássico da neuropsicologia cognitiva. Já quem concebe a psiquiatria como integrada à neurociência cognitiva e como um ramo da mesma insiste que a construção de tais modelos deve ser sensível ao conhecimento dos mecanismos neurais, já que são eles que processam as informações por meio de representações, possibilitando assim o comportamento inteligente do ser humano.

O que se está em questão aqui é uma tese ontológica sobre o mental, que implica uma determinada estratégia explicativa coerente e uma compreensão específica do papel dos modelos cognitivos em tal empreitada. A ideia de fundo é que o trabalho da neurociência cognitiva é basicamente descrever o funcionamento de mecanismos neurocognitivos que realizam as operações cognitivas. Ao fim e a cabo, estas são os *explanandum*, isto é, os fenómenos (memória, metarrepresentação, produção e compreensão de linguagem) que se constituem como objeto de teorização. Como os processos cognitivos são implementados por mecanismos cerebrais específicos, compreende-se que os modelos cognitivos devam ser suplementados pela descrição dos mecanismos neurocognitivos que os implementam, para que eles possam, de fato, desempenhar alguma função explicativa (GERRANS, 2014). Um modelo de reconhecimento de faces pode explicar o funcionamento geral desse processo num robô ou num ser humano. Nesse sentido, caso tal modelo capture propriedades e características importantes do sistema de reconhecimento facial humano, ele vai ser útil no desenvolvimento de uma teoria neurocognitiva que descreva o mecanismo neurocognitivo real responsável por tal processo. No decorrer da investigação, tal modelo pode

ser alterado significativamente, mas o decisivo é que ele ofereça um quadro conceitual inicial para a pesquisa simultânea dos diferentes níveis de funcionamento do respectivo mecanismo (nível molecular, intercelular, anatômico, neurofisiológico, redes neurais amplas, cognitivo), que possa gerar uma teoria unificada do reconhecimento de faces. A aplicação desse quadro conceitual para o domínio da psiquiatria implica, portanto, a busca pelas falhas e problemas nesses mecanismos neurocognitivos que geram os problemas psiquiátricos. A “psiquiatria neurocognitiva” amplia bastante o escopo da NC tal como estabelecido inicialmente.

Como se fez referência ao longo deste capítulo à ideia de *mecanismo* em neurociência cognitiva e ao respectivo modelo de explicação *mecanístico*, é oportuno algumas considerações gerais sobre tais conceitos. Dada a constatação de que tal modelo de explicação é mais adequado no contexto das ciências da vida do que o modelo nomológico-dedutivo importada da física, uma série de autores vem articulando um conjunto de teses de natureza ontológica e epistemológica e defendendo tal modelo no contexto da neurociência cognitiva e das ciências cognitivas de maneira mais ampla (CRAVER, 2001; 2007; BECHTEL & ABRAHAMSEN, 2005; BECHTEL & WRIGHT, 2009). Como destaca Leite (2018), a defesa de dessa estratégia explicativa é acompanhada de uma teoria mecanística da cognição humana que, por sua vez, endossa certa perspectiva sobre a relação entre o neural e a cognição. Um mecanismo, para dizer em poucas palavras, é um sistema com partes identificáveis organizado de uma tal maneira que as operações coordenadas realizadas pelas partes constituintes produzem uma determinada função (BECHTEL, 2008; CRAVER, 2007). A explicação mecanística procede justamente identificando essas partes e as operações realizadas através de suas interações, que geram o fenômeno alvo da investigação. É uma espécie de projeto de engenharia reversa (LEITE, 2018) que visa decompor o mecanismo e localizar (no cérebro, no caso das neurociências) as partes ativas, para que seja possível compreender de que modo elas produzem as operações constituintes dos processos cognitivos.

Há incontáveis funções realizadas por diferentes mecanismos neurais. Algumas dessas funções, como o controle neural do sistema respiratório, não são propriamente cognitivas. Já os mecanismos que geram os processos cognitivos caracterizam-se mais precisamente como mecanismos neurocognitivos ou, segundo a definição de Bechtel (2008), como mecanismos de processamento de informação. Estes possuem propriedades específicas que lhes possibilitam processar computacionalmente informações e, assim, instanciar os diferentes processos cognitivos humanos. Nesse arcabouço mecanístico, o neural não causa o mental no sentido de que eventos neurais antecedem no tempo e no espaço os eventos psicológicos; na verdade, há uma relação de identidade entre os mecanismos cognitivos e os mecanismos neurais. Mas, ao contrário de teoria da identidade clássica, que propõe uma identidade lógica entre processos e funções cognitivas e neurais, trata-se aqui de uma identidade heurística, segundo a formulação de Bechtel & McCauley (1999). Essa relação de identidade entre o cognitivo e o neural, assim compreendida, funciona como uma hipótese de trabalho para orientar a investigação científica e ajudar na permuta de métodos, análises e insights teóricos entre investigações situadas em níveis de análises diferentes.

A natureza hierárquica e estruturada em termos de múltiplos níveis de organização e complexidade é outra característica fundamental desses mecanismos. Isso, por sua vez, garante que a investigação dos fenômenos mentais possa se dar simultaneamente em termos de processamento de informação (nível cognitivo) e de processos e estruturas neurais, pois, ao fim ao cabo, são apenas diferentes maneiras de abordar e descrever os mesmos mecanismos. Além disso, essa atitude possibilita uma estratégia explicativa que não subordine as explicações cognitivas/psicológicas às explicações das neurociências. Apesar do nível cognitivo não ser ontologicamente autônomo em relação ao neural, a autonomia explicativa cognitiva é reconhecida, já que há processos autônomos de causação no interior de cada nível.



## Conclusão

Há, com efeito, muitos outros detalhes a serem especificados e discutidos no que se refere à tentativa de orientar a investigação e fundamentar a neurociência cognitiva de acordo com essa filosofia mecanística. Tentamos, por ora, apresentar apenas os seus elementos mais centrais, a partir da síntese de Leite (2018), para esclarecer de que forma a adoção desse aparato conceitual pode ser fértil para uma abordagem científica dos processos psicológicos e de suas disfunções.

A incorporação desse *background* conceitual da neurociência cognitiva no domínio da psiquiatria não está isenta de objeções e problemas, obviamente. Amplos setores das ciências cognitivas criticam as teses centrais desse quadro conceitual mecanístico, que incorpora conceitos centrais da teoria computacional da mente humana e explora seu potencial para solucionar os desafios filosóficos e empíricos de explicar a cognição humana e sua relação com o cérebro<sup>12</sup>. Porém, dada a amplitude do projeto e a complexidade das questões que ele tenta abordar, não é de se esperar outro cenário senão este em que as resoluções são mais provisórias e sujeitas a mais críticas e problemas do que se gostaria de admitir. Feitas tais ressalvas, no entanto, acreditamos que, caso a psiquiatria queira realmente se orientar em direção a uma abordagem explicativa dos fenômenos psicopatológicos, a integração com a neurociências cognitiva se apresenta como uma opção promissora. Isso se deve justamente à oportuna aliança entre investigação psicológica e investigação neural, tão cara a uma disciplina que, como ciência médica, não quer abrir mão do pressuposto materialista, ao mesmo tempo em que necessita de uma linguagem e de um vocabulário psicológico preciso para explicar a multiforme sintomatologia psiquiátrica.

---

<sup>12</sup> Nesse caso, temos em vista as abordagens enativas em ciências cognitivas que criticam elementos centrais da teoria computacional da mente, juntamente com ideia de que a cognição esteja restrita aos limites do cérebro e do corpo.

## Referências

- ÁGUIAR, M. P. Ascensão e consolidação da psiquiatria biológica norte americana: Uma análise histórica. **Mnemosini**, v. 11, p. 227-57, 2015.
- ANDREASEN, N. **The broken brain**: the biological revolution in psychiatry. New York: Harper & Row, 1984.
- \_\_\_\_\_. **Brave new brain**: conquering mental illness in the era of the genome. New York: Oxford University Press, 2001.
- BANZATO, C. & ZORZANELLI, R. Superando a falsa dicotomia entre natureza e construção social: o caso dos transtornos mentais. **Revista Latinoamericana de Psicopatologia Fundamental**, v. 17, n. 1, p. 100-13, Mar., 2014.
- BARON-COHEN, S.; LESLIE, A.; FRITH. Mechanical, behavioural and intentional understanding of picture stories in autistic children. **British Journal of Developmental Psychology**, v. 4, p. 113-25, 1986.
- BARON-COHEN, S. **Mindblindness**: An essay on autism and theory of mind. Boston: Bradford Book, 1997.
- BECHTEL, W. & MCCAULEY, R. Heuristic identity theory (or back to the future): the mind-body problem against the background of research strategies in cognitive neuroscience. In: HAHN, M. & STONESS, S. (Eds.). **Proceedings of the 21st annual meeting of the Cognitive Science Society**. Mahwah, (New Jersey): Lawrence Erlbaum Associates 1999, 67-72.
- BECHTEL, W. & ABRAHAMSEN, A. Explanation: a mechanist alternative. **Studies in History and Philosophy of Biological and Biomedical Sciences**, v. 36, n. 2, p. 421-41, 2005. DOI: 10.1016/j.shpsc.2005.03.010
- BECHTEL, W. **Mental mechanisms**: philosophical perspectives on cognitive neurosciences. New York: Routledge, 2008.
- BECHTEL, W. & WRIGHT, C. What is psychological explanation? In: Symons, J.; Calvo, P. (Eds.). **The Routledge companion to philosophy of psychology**. New York: Routledge/Taylor & Francis Group, 2009, p. 113-30.

BENTALL, R. **Madness explained**. London: Penguin, 2005.

\_\_\_\_\_. Understanding psychotic symptoms: cognitive and integrative models. In READ, J.; DILLON, J. (Eds.). **Models of madness: psychological, and biological approaches to psychosis**. Longres: Taylor & Francis Group, 2013, p. 270-90.

CARLAT, D. **Unhinged**: The trouble with psychiatry - a doctor's revelations about a profession in crisis. Free Press, 2010.

CASTAÑON, G.; JUSTI, F. & ARAÚJO, S. Cognitivismo. In: ARAÚJO, S; CAROPRESO, F; CASTAÑON, G & SIMANKE, R (Eds.). **Fundamentos filosóficos da psicologia contemporânea**. Juiz de Fora: ED. UFJF, 2014, p. 227-276.

CHURCHLAND, P. M. **The engine of reason, the seat of the soul**: A Philosophical Journey into the Brain. Cambridge: The MIT Press, 1995.

COLTHEART, M. Lessons from cognitive neuropsychology for cognitive science: a reply to Patterson and Plaut. **Topics in Cognitive Science**, v. 2, n.1, p. 3-11, 2009. Doi: 10.1111/j.1756-8765.2009.01067.x.

\_\_\_\_\_. Cognitive Neuropsychology. **Scholarpedia**, v 3, n. 2:3644, 2008. Disponível em: [http://www.scholarpedia.org/article/Cognitive\\_neuropsychology](http://www.scholarpedia.org/article/Cognitive_neuropsychology). Acesso em: 04 mar. 2020.

CRATSLEY, K. & SAMUELS, R. Cognitive science and explanation of psychopathology. In Fulford et al (Eds.). **Oxford handbook of philosophy and psychiatry**. Oxford, UK: Oxford University Press, 2013, p. 413-33.

CRAVER, C. Role functions, mechanisms, and hierarchy. **Philosophy of Science**, v. 68, p. 53-74, 2001. Doi: 10.1086/392866

\_\_\_\_\_. **Explaining the brain**: mechanisms and the mosaic unity of neuroscience. Oxford: Oxford University Press, 2007.

DAVID, A. Cognitive neuropsychiatry? **Psychological Medicine**, v. 23, n. 1, p. 1-5, 1993.

DAVID, A. & HALLIGAN, P. Editorial: Cognitive neuropsychiatry. **Cognitive Neuropsychiatry**, v. 1, p. 1-3, 1996.

- EY, H. Le développement mécaniciste de la psychiatrie: à l'abri du dualisme cartésien. In: \_\_\_\_\_. **Études psychiatriques I**. Paris: Desclée de Brouwer, 2006a, [1948], p. 51-66.
- \_\_\_\_\_. La position de la psychiatrie dans le cadres des sciences médicales. In: \_\_\_\_\_. **Études psychiatriques I**. Paris: Desclée de Brouwer, 2006b, [1948], p, 67-82.
- EYSENCK, M. & KEANE, M. **Manual de psicologia cognitiva**. 7. ed. Porto Alegre: Artmed, 2017.
- FODOR, J. **Psychological explanation: an introduction to the philosophy of psychology**. New York: Random House, 1968.
- FRITH, C. **The cognitive neuropsychology of schizofrenia**. Hove, UK: Lawrence Erlbaum, 1992.
- \_\_\_\_\_. Introduction to the classic edition. In: **The cognitive neuropsychology of schizofrenia**. New York: Psychological Press, p. XVI-XXIV, 2015.
- GARNHAN, A. Cognitivism. In: SYMONS, J. & CALVO, P. (Eds). **The Routledge companion to philosophy of psychology**. London: Routledge, 2009, p. 99-110.
- GERRANS, P. **The measure of madness: philosophy of mind, cognitive neuroscience, and delusional thought**. Cambridge: MIT Press, 2014.
- GUZE, S. **Why psychiatry is a branch of medicine**. New York: Oxford University Press, 1992.
- HALLIGAN, P. W. & MARSHALL, J. Towards a cognitive neuropsychiatry. In \_\_\_\_\_. (Eds.). **Method in madness: case studies in cognitive neuropsychiatry**. Hove: Psychology Press, 1996, p. 3-12.
- HALLIGAN, P. & DAVID, A. Cognitive neuropsychiatry: towards a scientific psychopathology. **Natural Reviews Neuroscience**, v. 2, n. 3, p. 209-15, 2001. Doi: 10.1038/3505858.
- HARRINGTON, A. **Mind fixers: psychiatry's troubled search for the biology of mental illness**. New York: W. W. Norton & Company, 2019.

HARRINGTON, L.; SIEGERT, R. & MCCLURE, J. Theory of mind in schizophrenia: a critical review. **Cognitive Neuropsychiatry**, v. 10, n. 4, p. 249-86, 2005. Doi: 10.1080/13546800444000056

HEALY, D. **The creation of psychopharmacology**. Cambridge: Harvard University Press, 2002.

INSEL, T. Transforming Diagnosis. **NIMH Director's Blog**. Bethesda, 29 April 2013. Disponível em: <https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml> >. Acesso em: 09 jun. 2020.

LEITE, D. **The twenty-first century mechanistic theory of human cognition**: a critical appraisal. Doctoral Dissertation - Department of Psychology and Cognitive Sciences of University of Trento. Trento, 2018.

KHALIDI, M. A. Interactive kinds. **British Journal for the Philosophy of Science**, n. 61, v. 2, p. 335-60, 2010.

KIRSCH, I. **The emperor's new drugs**: exploding the antidepressant myth. London: Bodley Head, 2009.

MAYES, R. & HORWITZ, A. V. DSM-III and the revolution in the classification of mental illness. **Journal of the History of Behavioral Sciences**, v. 41, n. 3, p. 249-67, 2005. Doi:10.1002/jhbs.20103

MURPHY, D. Explanation in psychiatry. **Philosophy Compass**, v. 5, n.7, p. 602-10, Doi: 2010. 10.1111/j.1747-9991.2010.00304.x

MURPHY, D. **Psychiatry in the scientific image**. Cambridge: MIT Press, 2006.

PORTER, R & MICALE, R. S. Reflexions on psychiatry and its histories. In: PORTER, R & MICALE R. S (Eds). **Discovering the history of psychiatry**. New York: Oxford Universit Press, 1994, p. 3-36.

PUTNAM, H. Minds and machines. In: Hook, S. (Ed.). **Dimensions of mind**. New York: Collier, 1961, p. 221-31.

SCHATER, D. **Searching for memory**: the brain, the mind, and the past. New York: Basic Books, 2008.

SHORTER, E. **A history of psychiatry: the era of asylum to the age of Prozac.** New York: John Wiley & sons, 1997.

\_\_\_\_\_. **A historical dictionary of psychiatry.** Oxford: Oxford University Press, 2005.

SIMANKE, R. **Metapsicologia lacaniana: os Anos de Formação.** Curitiba: Editora UFPR, 2002.

THAGARD, P. Cognitive architectures. In: FRANKISH, K & RAMSEY, W (Eds.). **The Cambridge handbook of cognitive science.** Cambridge: Cambridge University Press, 2012, p. 50-70.

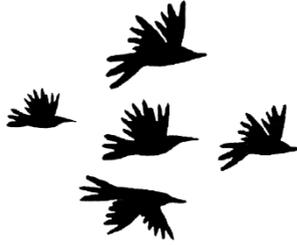
WALTER, H. The third wave of biological psychiatry. **Frontiers in Psychology**, v. 4, n. 582, p. 1-8, 2013. Doi:10.3389/fpsyg.2013.00582

WHITAKER, R. **Anatomy of an epidemic: magic bullets, psychiatric drugs, and the astonishing rise of mental illness in America.** New York: Crown Publishing, 2010.

WRIGHT, C. & BECHTEL, W. Mechanisms and psychological explanation. In: Paul Thagard (Ed.). **Philosophy of psychology and cognitive science.** Amsterdam: Elsevier, 2007, p. 31-79.

ZACHAR, P. **Psychological concepts and biological psychiatry: a philosophical analysis.** Amsterdam: John Benjamins Publishing Company, 2000.

\_\_\_\_\_. Psychiatric disorders: natural kinds made by the world or practical kinds made by us? **World Psychiatry**, v. 14, n. 3, p. 288-90, 2015. Doi: 10.1002/wps.20240



## Linguagem e psicologia em Ludwig Wittgenstein

*Thiago Ferreira dos Santos*<sup>1</sup>

*Os conceitos da psicologia são simplesmente conceitos do dia-a-dia. Não conceitos reformados pela ciência para seus próprios propósitos.*

**Wittgenstein**

Os últimos cursos ministrados por Wittgenstein como professor da Universidade de Cambridge, entre os anos de 1946 e 1947, foram na área de filosofia da psicologia; além disso, entre abril de 1946 e março de 1949, o filósofo dedicou seu trabalho à investigação de questões dessa natureza, levando-o a produzir um largo material a cerca dessa temática<sup>2</sup>. Foram escritas 1900 páginas de anotações que compõem os MSS 130-38. Posteriormente, seleções desse manuscrito foram feitas, não sendo, necessariamente, reordenadas por temas específicos internos a discussão sobre a psicologia senão em ordem cronológica dos manuscritos originais. Em dois momentos, Wittgenstein ditou os manuscritos a um datilógrafo, assim, no ano de 1947, uma seleção desse material foi datilografada no TS 229, que viria a ser as *Observações sobre a filosofia da psicologia I*, no ano

---

<sup>1</sup> Mestre em filosofia pela Universidade Federal de São Paulo (UNIFESP) e Professor de Filosofia do Instituto Federal de Alagoas (IFAL, *Campus* Penedo).

<sup>2</sup> Cf. P. M. S. Hacker (2010).

seguinte, em 1948, foi datilografado no TS 232, que viria a ser as *Observações sobre a filosofia da psicologia II*<sup>3</sup>.

Ainda da série de anotações que formam os MSS 130-38, a segunda parte do MS137 e o MS138 por inteiro, escritos entre 1948-49, foram publicados pelos editores como os *Últimos escritos sobre a filosofia da psicologia*. Em 1949, compilou uma seleção de anotações manuscritas no MS144 (cujo datiloscrito está perdido), que deu origem a *segunda parte das Investigações Filosóficas*, depois da 4<sup>o</sup> edição da obra ficou conhecida como *filosofia da psicologia – Um fragmento*.

Pelo caráter cronológico, de forma e conteúdo das observações apresentadas por Wittgenstein, de acordo com Anscombe<sup>4</sup>, o teor dos manuscritos MSS 130-38 podem ser caracterizados como estudos preliminares para esta *filosofia da psicologia – Um fragmento*.

Entendemos que entre as décadas de 30 e 40, Wittgenstein desenvolveu um trabalho vasto e importante que o levou a suas concepções maduras apresentadas nas *Investigações filosóficas*. E já em seus 693 parágrafos apresentados pelos editores como a parte I de suas *Investigações*, encontramos uma série de exemplos que nos apresentam questões relacionadas à significação de conceitos psicológicos, e que serão tratados de forma mais específica nos escritos posteriores a 1946.

These controversial, indeed revolutionary, conceptual commitments are prominent in the *Investigations*. They are all grammatical clarifications supported by reasoned argument. Their denial leads to incoherence. And they provide a very substantial grammatical framework for more detailed investigations of the large network of psychological concepts that inform the lives of language-using creatures like us. It was to such investigations that Wittgenstein turned in 1946 (HACKER, 2010, p. 10).

Autores como Hacker entendem que, ainda que importantes de serem trazidas nas incursões de Wittgenstein sobre a filosofia da

---

<sup>3</sup> Cf. L. Wittgenstein (1980).

<sup>4</sup> Cf. Prefácio do livro **Remarks on the Philosophy of Psychology 2 vols.** Oxford Basil Blackwell, 1980.

psicologia, as questões estreitamente conectadas ao problema da linguagem e significado já estão bem trabalhadas nas *Investigações filosóficas*, de modo que nos últimos escritos se situam como um pano de fundo mais amplo das discussões que sugerem um tratamento mais específico. Em outras palavras, o uso significativo da linguagem é pressuposto fundamental no trabalho com a filosofia da psicologia de seus últimos escritos, que são resultado de uma inclinação que parte de suas investigações do significado e o levam na década de 40 cada mais fundo em questões concernentes à psicologia.

Como sabemos o trabalho com a linguagem desde aquele apresentado no *Tractatus*, como sua concepção madura apresentada nas *Investigações* sofreram um trabalho de leitura e interpretação bastante vasto. Nos anos que seguiram a morte de Wittgenstein e a publicação das *Investigações filosóficas em 1953*, houve uma série de estudos da obra do filósofo tomando como objeto de investigação a questão do significado, que se pode dizer, ainda atualmente é a temática mais amplamente discutida na obra do filósofo. Em contra partida, quando se trata de sua filosofia da psicologia, os estudos mais relevantes aparecem mais recentemente. Somente nos anos 80, em média três décadas depois das anotações originais, o restante do material que versa sobre questões concernentes a conceitos psicológicos é publicado, trazendo consigo uma gama de investigações em que Wittgenstein inova intensamente o tratamento das questões relacionadas à psicologia oferecendo novas perspectivas na compreensão de conceitos psicológicos.

### **Filosofia da psicologia como investigação conceitual**

Quando se observa o tratamento de Wittgenstein dos conceitos psicológicos é preciso ter em mente que a psicologia se desenvolve no final do século XIX e início do século XX como ciência de caráter empírico e experimental. É certo que, no período anterior ao final do século XIX, não se pode falar em pesquisas sistemáticas para se tratar fatos psicológicos.

Com o advento dos ideais positivistas e das propostas científicas, estabeleceu-se a exigência de mensuração, de quantificação e de método aplicado às ciências naturais na produção de conhecimento; com o intuito de abranger a totalidade da investigação, reivindicou-se também que a psicologia (que pesquisa os processos interiores) tomasse para si os pressupostos das ciências que estavam se desenvolvendo neste período.

Com o avanço tecnológico, novas respostas e novas perguntas para o tema foram sendo cada vez mais frequentes e assim novos conhecimentos foram se formando, sobretudo de cunho físico-biológico, tendo como principal vertente de pesquisa a psicofísica que pretendia, grosso modo, traçar um paralelo entre mente e cérebro, isto é, se propunha a estabelecer o estudo do *interior* com as mesmas ferramentas de investigação das ciências naturais para o mundo exterior, tratando, assim, essas duas instâncias de forma similar e, conseqüentemente, constituindo certo estudo empírico do *interior*, o que se pode chamar um determinismo psicofísico em que, aparentemente, seria possível prever, mensurar, quantificar os estados internos. Porém, o trabalho de Wittgenstein com a psicologia não se insere nessa perspectiva, não está no âmbito da experimentação.

A pesquisa de Wittgenstein não se situa em uma perspectiva empírica, experimental, mas explicita uma investigação de cunho conceitual, de esclarecimento dos conceitos da psicologia. Tendo em vista que “os conceitos da psicologia são simplesmente conceitos do dia-a-dia. Não conceitos reformados pela ciência para seus próprios propósitos, como os da física e da química” (WITTGENSTEIN, 2008, p. 273; §62). Mas se situam como parte de nossas práticas.

Nesse contexto, a partir de suas incursões na filosofia, Wittgenstein trabalhou intensamente apresentando suas concepções sobre a metafísica e o processo de dissociação da mesma. Clareando as confusões presentes em concepções que combinam questões empíricas com questões conceituais; tendo como problema geral a questão da linguagem que aparece como fonte dessas confusões. Nesse aspecto, o estudo dos

conceitos psicológicos em Wittgenstein não é pautado na análise do fenômeno empírico, fisiológico, mas no âmbito da investigação conceitual. Nos termos de Wittgenstein:

Investigações filosóficas: investigações conceituais. O que é essencial à metafísica: que para ela não seja clara a distinção entre investigações factuais e conceituais. A questão metafísica sempre tem toda a aparência de uma questão factual, embora o problema seja conceitual (WITTGENSTEIN, 2010, p. 202; §949).

Wittgenstein dá ênfase a esse problema, pois entende que vivências em primeira pessoa não são explicadas factualmente. E como proposto por Hacker, Wittgenstein prestou mais atenção do que qualquer outro filósofo a questão da assimetria entre primeira e terceira pessoa do presente em proposições da psicologia. De modo que, no movimento de passagem para a psicologia nas *Investigações filosóficas*, essa é, em última instância, a temática mais fortemente trabalhada pelo filósofo no clareamento e na compreensão do uso de nosso vocabulário psicológico e que nos apresenta o modo como Wittgenstein entende a psicologia já nas *Investigações*. E para isso, há um desmonte da concepção metafísica da linguagem e do interior que sustentam a assimetria.

Durante séculos, as concepções de interior se sustentaram em teorias *dualistas* e *monistas*. As primeiras afirmam que o mundo é formado por substâncias materiais e imateriais que são independentes e irreduzíveis entre si (*e.g.*, dualismo da substância), neste contexto, substâncias físicas e mentais são inconciliáveis; as segundas afirmam que existe apenas uma substância – material – responsável tanto pelo físico quanto pelo mental, isto é, o campo mental só é possível devido ao físico e as experiências mentais se reduzem às físicas. Todavia, tanto para concepções dualistas, como para concepções monistas, o interior foi visto como um âmbito privado. O plano do mental é visto por grande parcela da tradição filosófica como algo privado, isto é, as experiências mentais ocorrem e são

conhecidas apenas pelo sujeito da experiência, em outras palavras, o sujeito teria posse e conhecimento privado de seu próprio caso.

Podemos comparar essa noção de interior com a ideia de um teatro mental, em que as experiências internas ocorrem em um palco do qual só o sujeito da experiência é expectador, somente ele tem acesso ao conhecimento interno. Sobre esta concepção de interior, pode-se dizer, nesta noção de teatro mental, está implícito o *modelo agostiniano de linguagem*<sup>5</sup>, pois se aplica o modelo objeto-designação, tornando o tratamento dado ao mental semelhante ao dado ao material (GLOCK, 1998, p. 218). Em outras palavras, se explicita na noção de teatro mental o pilar da crítica wittgensteiniana, que é a influência da linguagem agostiniana (objeto - designação) na concepção de interior e, conseqüentemente, no modo como são tratados os conceitos da psicologia.

A caracterização de um conceito psicológico seja pela filosofia, seja pela psicologia - e mesmo no nosso uso cotidiano - acompanha, em sua definição, uma imagem de linguagem, e o uso desta imagem de linguagem carrega uma série de conseqüências que imprime naquele modelo suas características. A *imagem agostiniana da linguagem* que é fundamentalmente referencialista, isto é, supõe-se que, de modo geral, as palavras nomeiam objetos ou descrevem objetos, o significado de uma palavra seria dado pela sua referência. E supondo que esta imagem se aplica a todos os conceitos para o estabelecimento de significado, quando esta imagem é utilizada para representar conceitos psicológicos, ela reaparece como descrição de um ambiente interno. Isto é, temos a descrição de conceitos que se referem ao ambiente externo, um ambiente observável, público e a descrição de conceitos que se referem ao ambiente interno, um ambiente privado, acessível apenas ao sujeito da experiência, com o uso de uma linguagem que referiria a objetos privados, processos privados etc. (SANTOS, 2017, p. 13), a ideia do interior como teatro mental.

Neste sentido, o plano interior seria uma instância acessível apenas para o sujeito da experiência, seria experienciado, conhecido diretamente,

---

<sup>5</sup> Cf. *Investigações Filosóficas*, §1.

apenas por ele próprio, sendo explicado a outro sujeito apenas indiretamente, este outro não teria acesso ao conhecimento direto, real, mas apenas ao comportamento, pois não teria acesso a outro interior se não o seu próprio. Tem-se então uma privacidade epistêmica da experiência, e.g., “somente eu conheço minhas dores”. Isto gera o que chamamos de assimetria pronominal entre 1ª e 3ª pessoa, em que a 1ª pessoa tem primazia do conhecimento interno próprio, em detrimento da 3ª pessoa que só teria acesso indireto, por meio de observação exterior.

A resposta de Wittgenstein a questão da “posse e privacidade epistêmica” da experiência interna (assimetria pronominal) é que este não é um problema ontológico ou epistemológico, mas que se esclarece como um problema da linguagem (SANTOS, 2017, p. 14). E a partir de uma crítica da linguagem é que Wittgenstein dá o tom de seu tratamento do ambiente dos conceitos psicológicos.

No §244 das *Investigações Filosóficas* é dado ênfase de que a exteriorização<sup>6</sup> desse plano interior é linguística, estando relacionada a um sistema linguístico primitivo que intersubjetivamente é substituído por sistemas linguísticos mais complexos possibilitados pelas formas de vida no qual estão inseridos os sujeitos da experiência. Assim, o filósofo afirma:

How do words *refer* to sensations? – There doesn't seem to be any problem here; don't we talk about sensations every day, and name them? But how is the connection between the name and the thing named set up? This question is the same as: How does a human being learn the meaning of names of

---

<sup>6</sup> Esse termo foi introduzido na filosofia por Ryle, mas é também comum a sua utilização para traduzir os termos wittgensteinianos *Ausserung e Ausdruck* (outras alternativas são “expressão”, “manifestação” ou “proferimento”). Wittgenstein caracterizava como exteriorizações alguns usos de sentenças psicológicas construídas em primeira pessoa e no tempo presente. Negativamente, isso indica que não se trata de descrições ou relatos de entidades mentais privadas, encontradas em um domínio interno. Positivamente, Wittgenstein caracteriza as exteriorizações como sendo dotadas do mesmo tipo de carga expressiva de um gesto ou de um franzir de testa, capazes de expressar ou manifestar emoções, atitudes etc. São substitutos parciais para expressões naturais da esfera mental, tais como gritos, sorrisos ou caretas, constituindo extensões aprendidas destas expressões. As palavras designadoras de sensações “ligam-se às expressões primitivas, naturais, de sensação, sendo utilizadas em seu lugar. Uma criança se machuca e chora; adultos então falam com ela, ensinando-lhe exclamações e, depois, sentenças. Ensinam à criança um novo comportamento de dor” (PI §244). Não se trata aqui de teorizações superficiais sobre o processo-aprendizagem, mas antes da afirmação de que, do ponto de vista lógico, a função das exteriorizações é semelhante à das manifestações não verbais. Cf. GLOCK (1998, p. 153-4).

sensations? For example, of the word “pain”. Here is one possibility: words are connected with the primitive, natural, expressions of sensation and used in their place. A child has hurt himself and he cries; then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour.

“So you are saying that the word ‘pain’ really means crying?” – On the contrary: the verbal expression of pain replaces crying, it does not describe it (WITTGENSTEIN, 2009, p. 95; §244).

A relação entre interior e exterior se dá justamente por exteriorização que imprime nessa relação como marca a *expressão linguística*, ao passo que, *toda linguagem é pública*.

### **Argumento da linguagem privada como aspecto fundamental da investigação**

Wittgenstein recusa a ideia de uma linguagem privada na significação de um conceito e reivindica o tratamento do ambiente dos conceitos psicológicos como possível somente a partir de uma linguagem pública. Suas reivindicações contra uma linguagem radicalmente privada aparecem nas *Investigações filosóficas* em um grupo de parágrafos conhecido largamente como argumento da linguagem privada. Este argumento, na verdade se situa em um contexto mais amplo de debates que remonta o percurso de dissociação levado a cabo por Wittgenstein desde os primeiros parágrafos das *Investigações filosóficas* com a recusa da “imagem Agostiniana de linguagem”. Os parágrafos em que aparece o argumento da linguagem privada, na verdade, explicitam uma discussão que já está presente na obra, isto é, compõe uma unidade. A passagem para a psicologia nessa obra é, portanto, um movimento natural no percurso de observações do filósofo quanto ao uso significativo da linguagem.

De acordo com Ter Hark:

In the Augustinian idea of language (see PI, § 1-27), defining by ostension is an important way of giving words meaning. Wittgenstein's objection is that

ostensive ceremonies can only be meaningfully conducted within horizontal relations of language games. An ostensive definition of 'ball' is only meaningful if it is already clear what can and cannot be done with a ball. The possibility of ostensive definitions is conditional on the use of the concept within a language-game. So the meaning of a concept is not its reference but its use (TER HARK, 1990, p. 94).

Em outras palavras, o argumento da linguagem privada se constrói, por um lado, como a recusa de uma concepção referencialista de linguagem e, por outro, da explicitação de que a significação de um conceito se dá em seu uso. Desse modo, no percurso do argumento há uma recusa de que processos internos, conceitos mentais, intencionais etc., signifiquem a partir de uma concepção referencialista de linguagem.

Nos §§243-315 das *Investigações Filosóficas* (WITTGENSTEIN, 1996, p. 98-110), Wittgenstein reivindica a publicidade da linguagem que estabelece um interior expresso no uso significativo da linguagem. No percurso do argumento, se explicita, em linhas gerais, que (i) o tratamento dos conceitos psicológicos não é privado; (ii) não é possível uma linguagem privada, a linguagem é pública e; (iii) com isto não se exclui o interior, mas se reconstrói uma concepção de “interior que se mostra nos conceitos que utilizamos para expressá-lo” (FATTURI, 2010, p. 18).

Wittgenstein inicia o argumento fazendo distinções quanto ao que ele está chamando de linguagem privada, para que não haja precipitações ou dúvidas quanto ao que está sendo criticado no argumento.

A discussão da linguagem privada se inicia por distinguir entre o sentido ordinário de privacidade, no qual meu diário oculto, meu código secreto ou minha dor escondida são questões privadas, e a concepção sublime de privacidade que é foco de interesse do interlocutor. Ainda que um diário, um código ou uma dor possam muito bem ser privados – ou seja, ninguém mais sabe a seu respeito – é sempre possível, pelo menos em princípio, que outros descubram algo sobre estas coisas. O interlocutor, por outro lado, nos pede para imaginar uma linguagem que é necessariamente privada, uma que ninguém mais teria a possibilidade de entender, pois as palavras “se referem à aquilo que só pode ser conhecido pela pessoa que fala; às suas sensações

privadas imediatas. Assim, uma outra pessoa não pode entender a linguagem” (§243) (STERN, 2009, p. 255).

No §243 explica que não se pretende falar de uma linguagem estabelecida em monólogos que um sujeito poderia fazer para si mesmo, ou até mesmo uma linguagem em códigos, pois é passível de compartilhar traduzindo-se para uma linguagem comum. A linguagem aqui tratada é aquela em que apenas o sujeito pode saber. Aquela pela qual é possível um acesso direto a suas experiências privadas imediatas, não sendo possível um compartilhamento. É este o modelo de linguagem privada a ser criticado pelo filósofo no decorrer dos parágrafos ulteriores. A seguinte passagem, a partir do questionamento do narrador e da resposta do interlocutor, expressa bem isso:

A human being can encourage himself, give himself orders, obey, blame and punish himself; he can ask himself a question and answer it. So one could imagine human beings who spoke only in monologue, who accompanied their activities by talking to themselves. – An explorer who watched them and listened to their talk might succeed in translating their language into ours. (This would enable him to predict these people’s actions correctly, for he also hears them making resolutions and decisions.)

But is it also conceivable that there be a language in which a person could write down or give voice to his inner experiences – his feelings, moods, and so on a for his own use? — Well, can’t we do so in our ordinary language? – But that is not what I mean. The |89| words of this language are to refer to what only the speaker can know – to his immediate private sensations. So another person cannot understand the language (WITTGENSTEIN, 2009, p. 95e; §243).

O argumento de Wittgenstein para demonstrar a impossibilidade de uma linguagem privada se constrói tendo ênfase, pelo menos, nos §§258, 265, 293. Um dos principais argumentos explicitados nos parágrafos citados é a recusa da ideia de ostensão interna, que reaparece nesse grupo de parágrafos como modelo de significação dado privadamente. Se por um lado, a partir dos primeiros parágrafos das *Investigações filosóficas* Wittgenstein já apresenta uma série de exemplos que levam a uma

dissociação do modelo ostensivo como fundante da linguagem, quando se trata de significação de objetos externos, nesses parágrafos que compõem o argumento da linguagem privada, o filósofo vai argumentar quanto à ostensão privada e ainda contra a possibilidade de um critério de uso privado.

O primeiro empecilho é a *impossibilidade de fazer uma definição ostensiva do interior por introspecção*, apresentando no §258, que situa o núcleo do argumento da linguagem privada. Nesse parágrafo, Wittgenstein propõe a composição de um diário para a ocorrência de determinada sensação:

Let's imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "S" and write this sign in a calendar for every day on which I have the sensation. -- I first want to observe that a definition of the sign cannot be formulated. a But all the same, I can give one to myself as a kind of ostensive definition! a How? Can I point to the sensation? a Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation a and so, as it were, point to it inwardly. a But what is this ceremony for? For that is all it seems to be! A definition serves to lay down the meaning of a sign, doesn't it? a Well, that is done precisely by concentrating my attention; for in this way I commit to memory the connection between the sign and the sensation. a But "I commit it to memory" can only mean: this process brings it about that I remember the connection *correctly* in the future. But in the present case, I have no criterion of correctness. One would like to say: whatever is going to seem correct to me is correct. And that only means that here we can't talk about 'correct' (WITTGENSTEIN, 2009, p. 98-99e; §258).

A impossibilidade de se definir ostensivamente o interior, por meio de introspecção, dá-se pela falta de um critério de correção de uso interior que justifique a definição. Sem um critério público de uso não há como falar de uso "correto" ou "incorreto", pois não é suficiente a própria atenção como critério próprio, ao passo que qualquer definição linguística só se justifica publicamente.

O segundo empecilho é o *problema de se recorrer à memória como um critério de uso privado*, apresentado no §265:

Let us imagine a table, something like a dictionary, that exists only in our imagination. A dictionary can be used to justify the translation of a word X by a word Y. But are we also to call it a justification if such a table is to be looked up only in the imagination? a “Well, yes; then it is a subjective justification.” a But justification consists in appealing to an independent authority a “But surely I can appeal from one memory to another. For example, I don’t know if I have remembered the time of departure of a train correctly, and to check it I call to mind how a page of the timetable looked. Isn’t this the same sort of case?” No; for this procedure must now actually call forth <sup>94</sup>the *correct* memory. If the mental image of the timetable could not itself be *tested* for correctness, how could it confirm the correctness of the first memory? (As if someone were to buy several copies of today’s morning paper to assure himself that what it said was true.)

Looking up a table in the imagination is no more looking up a table than the image of the result of an imagined experiment is the result of an experiment (WITTGENSTEIN, 2009, p. 100e).

A questão aqui remete ao mesmo problema apresentado no §258, que é a falta de um critério que justifique o uso e, recorrer à memória como um critério de correção de uso é confuso, justamente porque não faz sentido supor que a memória seja critério de uso dela própria<sup>7</sup>.

O terceiro empecilho contra a ideia de uma linguagem privada é quanto à *irrelevância da existência do objeto interno para a compreensão do sentido de um conceito psicológico*. Este argumento é importante para a compreensão do modo como Wittgenstein entende a dimensão de tratamento de conceitos psicológicos, quanto explicita a irrelevância na referência a um objeto interno, mas, em contrapartida, está preocupado em investigar o sentido desses conceitos a partir de um contexto de uso regrado. No §293, Wittgenstein propõe o seguinte:

---

<sup>7</sup> Cf. *Remarks on the Philosophy of Psychology* Vol. 1, §§ 903-8.

If I say of myself that it is only from my own case that I know what the word “pain” means a must I not say *that* of other people too? And how can I generalize the *one* case so irresponsibly? Well, everyone tells me that he knows what pain is only from his own case! — Suppose that everyone had a box with something in it which we call a “beetle”. No one can ever look into anyone else’s box, and everyone says he knows what a beetle is only by looking at *his* beetle. Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. But what if these people’s word “beetle” had a use nonetheless? If so, it would not be as the name of a thing. The thing in the box doesn’t belong to the language-game at all; not even as a *Something*: for the box might even be empty. No, one can ‘divide through’ by the thing in the box; it cancels out, whatever it is.

That is to say, if we construe the grammar of the expression of sensation on the model of ‘object and name’, the object drops out of consideration as irrelevant (WITTGENSTEIN, 2009, p. 106-7e).

Este empecilho resulta das impossibilidades apontadas pelos dois primeiros, visto que, por introspecção, não é possível definir um objeto interno justamente pela falta de critérios internos de correção que justifiquem isto, ao passo que, na linguagem pública para se compreender o uso de um conceito, tal como “besouro”, sequer é necessário apontar para algo dentro da caixa, como propõe o exemplo. Isto porque para se compreender o uso de um conceito psicológico, são necessários critérios de uso dados publicamente, e no jogo de linguagem de um conceito o que lhe dá sentido é seu uso. Deste modo, Wittgenstein reivindica a linguagem pública como única possibilidade de linguagem.

As diferenças pretendidas pelo filósofo não são diferenças ontológicas, não é uma dicotomia de mundos em sentido ontológico, mas são “diferenças *gramaticais*, e isto significa que palavras tais como *interno* e *privado*, fazem parte do jogo de linguagem dos termos psicológicos para marcar as diferenças de tais conceitos com os conceitos de outros jogos de linguagem” (DONAT, 2009, p. 452). Não estando, assim, a compreensão ligada a um conhecimento interno (dado por ostensão privada), mas a uma atividade baseada no uso. No jogo de linguagem das sensações, não

está a compreensão de um objeto interior, mas a expressão de nossas sensações em diversas manifestações a qual cabem os usos.

## Referências

- DONAT, Mirian. Wittgenstein e as supostas posse privada e privacidade epistêmica da experiência. **Revista de Filosofia Aurora**, vol 21, n. 29, jul./dez., p. 437-53, 2009.
- FATTURI, Arturo. **Mundo interior e expressão: a filosofia da psicologia de Ludwig Wittgenstein**. – Tese de Doutorado, São Carlos, 2010.
- GLOCK, Hans-Johann. **Dicionário Wittgenstein**. Trad. Helena Martins; rev. Técnica de Luiz Carlos Pereira. Rio de Janeiro: Jorge Zahar Ed., 1998.
- HACKER, P. M. S. The Development of Wittgenstein's Philosophy of Psychology. In: John Cottingham & Peter Hacker (eds.). **Mind, Method, and Morality: Essays in Honour of Anthony Kenny**. Oxford: Oxford University Press, 2010.
- \_\_\_\_\_. Some Remarks on Philosophy and on Wittgenstein's Conception of Philosophy and its Misinterpretation. **Argumenta**, vol. 1, n. 1, University of Sassary, p. 43-58, 2015.
- SANTOS, Thiago F. Memória e conteúdo na filosofia da psicologia de Ludwig Wittgenstein. **Estudos Filosóficos**, nº 16, p. 12-22, 2017.
- TER HARK, Michel. **Beyond the inner and the outer Wittgenstein's philosophy of psychology**. Synthese Library (Vol. 214), 1990.
- WITTGENSTEIN, Ludwig. **Investigações filosóficas**. Trad. José Carlos Bruni. São Paulo: Nova Cultural, 1996. [Col. Os Pensadores].
- \_\_\_\_\_. **Philosophische untersuchungen/ Philosophical investigations**. Translated by G.E.M. Anscombe, P. M. S. Hacker & Joachim Schulte. – Rev. 4th ed. / by P. M. S. Hacker & Joachim Schulte. Oxford: Wiley Blackwell, 2009.
- \_\_\_\_\_. **Observações sobre a filosofia da psicologia – vol I e II**. Trad. Ricardo Hermann Ploch Machado. São Paulo: Idéias & Letras, 2008. [Subjetividade Contemporânea].
- \_\_\_\_\_. **Remarks on the philosophy of psychology 2 vols**. Oxford: Basil Blackwell, 1980.
- \_\_\_\_\_. **Last writings on the philosophy of psychology 2 vols**. Oxford: Basil Blackwell, 1992.



## A descrição dos próprios estados mentais

*Christopher Alexander Franke*<sup>1</sup>

*Joelma Marques de Carvalho*<sup>2</sup>

O objetivo principal deste capítulo é analisar em que sentido podemos descrever os nossos próprios estados mentais especialmente a partir de uma interpretação feita por Hacker (2000) sobre as obras do segundo Wittgenstein e tendo como interlocutor o filósofo Franz Brentano em seu livro “Psicologia do Ponto de vista Empírico” (1874). Para isso, nós dividimos esse capítulo em 4 partes. Na primeira parte, trataremos de modo introdutório o jogo de linguagem dos conceitos mentais. Na segunda parte, abordaremos alguns significados de “descrição” por meio da análise linguística de vários proferimentos feitos na primeira pessoa do singular com termos mentais. Na terceira parte, discutiremos as razões apresentadas por Hacker para distinguir os modos de descrição feitos sobre objetos físicos e sobre os próprios estados psicológicos. Por último, apresentaremos alguns argumentos pelos quais consideramos a abordagem da descrição dos próprios estudos psicológicos a partir dessa

---

<sup>1</sup> Graduado e Mestre em Filosofia, Ciências Políticas e Economia (M.A.) pela Ludwig-Maximilians-Universität (LMU) e Doutor em Filosofia (Dr. phil.) pela mesma instituição.

<sup>2</sup> Doutora em Filosofia pela Ludwig-Maximilians-Universität (LMU) em Munique na Alemanha. Atualmente, é Pesquisadora Sênior no Curso de Filosofia na Universidade de Salzburgo na Áustria.

interpretação wittgensteiniana como sendo apenas parcialmente correta. Contudo, não discutiremos aqui se a interpretação desse especialista acerca da filosofia de Wittgenstein é correta ou não.

## 1. O jogo de linguagem dos conceitos mentais

Para Wittgenstein, as palavras e frases não possuem um significado determinado ou exato, pois o sentido dos termos que fazem parte de nossos proferimentos podem variar de acordo com o contexto ou com o *jogo de linguagem*. A expressão “jogo de linguagem” é introduzida por esse autor com o objetivo de apontar para o fato de que a linguagem não é algo fixo, mas um fenômeno social que está em constante mudança e faz parte de nossas formas de vida. Sendo assim, nós podemos entender a linguagem como um conjunto de diferentes jogos de linguagem, isto é, de diversas práticas de ensino, atividades linguísticas existentes e hipotéticas que também incluem várias atividades interligadas com a linguagem. Exemplos de jogos de linguagem são: fazer sugestões, descrever objetos, escrever ou ler uma história, contar piadas, apresentar os resultados de experimentos por meio de tabelas ou diagramas, cantar etc (WITTGENSTEIN, 1952, §23).

Essa visão da linguagem difere radicalmente do modelo tradicional defendido em sua obra *Tractatus Logico-Philosophicus* (2001), no qual cada palavra deve corresponder a um objeto existente no mundo para ter sentido. Além disso, a determinação do significado e da referência dos termos não são realizados privadamente, mas intersubjetivamente. A linguagem é assim um fenômeno social e público. Isso quer dizer que os significados das palavras não fazem parte de uma linguagem privada, inclusive os conceitos psicológicos ou mentais tais como “dor”, “raiva”, “saúde”, “pretender” etc.

O jogo de linguagem que nos interessa aqui é o dos conceitos mentais. Conforme Wittgenstein, o jogo de linguagem dos conceitos psicológicos é muito mais sutil do que se imagina e pode nos levar a defender ideias falsas

acerca da mente. A primeira ideia que devemos desmistificar é que os significados dos termos mentais são determinados de modo privado. Ao contrário, podemos saber o significado dessas palavras independentemente de “termos” a experiência dos estados mentais, aos quais essas expressões se referem. Nesse sentido, um homem pode usar o termo “cólicas menstruais” corretamente apesar de nunca ter experenciado cólicas menstruais.

Uma vez que vários problemas filosóficos são frutos de um mal entendimento do funcionamento da linguagem e dos diversos usos das palavras, é necessário fazer uma análise conceitual cuidadosa dos termos psicológicos de nossa linguagem para resolver ou dissolver esses problemas filosóficos. Esse método de descrição dos sentidos das palavras é o ponto de partida para entendermos a visão wittgensteiniana acerca da descrição dos fenômenos mentais. Por esse motivo, seguiremos esse método aqui e analisaremos sentenças com termos mentais na primeira pessoa do singular com o objetivo de avaliarmos em que sentido podemos descrever os nossos próprios fenômenos mentais.

## **2. Significados de “descrição”**

Assim como os outros termos de nossa linguagem, a palavra “descrever” pode ser usada de diferentes formas. Imaginemos a seguinte situação: Um sujeito A compra um celular x numa loja e o descreve para um amigo (sujeito B): O meu celular é cinza, pesa 192 gramas, possui uma memória de 8 GB Ram, altura de 158 mm etc. Nesse caso, podemos afirmar:

(1) “O sujeito A descreveu o seu novo celular para seu amigo B”

Na frase acima (1), o verbo “descrever” está interligado com outras atividades tais como ver, observar e examinar um objeto por meio da percepção visual. Além dessas práticas, o indivíduo A também pode buscar informações detalhadas sobre esse produto em um manual. Nesse âmbito,

um especialista em celulares pode descrever melhor o novo celular de A do que ele mesmo. Imaginemos agora que o sujeito A experencie tristeza constante e inquietação e, por isso, decide procurar um psiquiatra. Na clínica, o psiquiatra lhe pergunta qual é o motivo pelo qual ele lhe procurou e o que ele está sentindo. Como resposta, ele faz três declarações, a saber:

(2) “Eu tenho tristeza constante”

(3) “Eu tenho inquietação”

(4) “Por isso, eu acho que estou com depressão”

Nesse contexto, poderíamos afirmar acerca do sujeito A:

(5) “O sujeito A descreveu seus estados mentais para o psiquiatra”

O verbo “descrever” no enunciado (5) acima seria um relato acerca de alguns estados mentais que frequentemente são associados ao cérebro de A. Devemos observar aqui que a expressão “estados mentais” nessa sentença se refere aos fenômenos psicológicos em geral tais como tristeza, inquietação, insônia e pensamentos. Não existe uma definição exata para “estados mentais”, mas geralmente usamos essa expressão para nos referir aos fenômenos que ocorrem na mente, os quais podem ser conscientes ou inconscientes. Porém, conforme a interpretação de Hacker (2000, 41) acerca da obra *Investigações Filosóficas*, Wittgenstein usa “estados mentais” em um sentido estrito e se refere apenas aos fenômenos psicológicos conscientes, que possuem uma duração genuína, isto é, que podem ser suspensos durante o sono e retomados posteriormente. Geralmente esses termos são usados no gerúndio e se referem a ações ou reações de um sujeito. Nesse sentido, ter a intenção de algo, pensar e acreditar não são considerados como sendo estados mentais. Infelizmente, Hacker não oferece um exemplo claro de “estados mentais” nesse sentido. Mesmo assim, podemos afirmar que exemplos de “estados mentais” em Wittgenstein são: estar com raiva, sentir uma dor etc.

Se essa interpretação de Hacker estiver correta, então a sentença (4) não pode ser uma descrição de um estado mental, uma vez que a atividade de pensar não é um estado psicológico. Ao enunciar (4), o indivíduo A

informa apenas que não está seguro acerca do conteúdo do pensamento expresso. Contudo, estar triste e inquieto podem ser considerados como “estados mentais” no sentido estrito. Sendo assim, resta-nos analisar em que sentido podemos afirmar que os enunciados (2) e (3) são descrições de alguns estados mentais de A, ou seja, que a afirmação (5) faz sentido dentro desse contexto.

Para isso, devemos refletir acerca dos possíveis significados do verbo “descrever” na segunda situação e fazer uma comparação entre o uso de “descrever” nas sentenças (1) e (5). Na frase (5), o verbo “descrever” parece não estar interligado às mesmas atividades de (1), visto que nesse caso não podemos falar de uma observação por meio da percepção visual. O sujeito A não pode, por exemplo, verificar os seus estados mentais de modo mais aproximado, pois não existe um olho interno através do qual é possível examiná-los. Além disso, devemos observar que o termo “ter” nas declarações (2) e (3) não têm o sentido de posse. Imaginemos que o sujeito A faça a seguinte afirmação:

(6) “Eu tenho um novo celular”

No enunciado acima, podemos afirmar que o verbo “ter” tem o sentido de posse. Por esse motivo, o sujeito A pode emprestar esse objeto ou vendê-lo. Os estados mentais não são objetos análogos a celulares, os quais podem ser possuídos, emprestados ou vendidos. A rigor, só podemos falar de uma posse dos estados mentais em um sentido metafórico. Em nossa opinião, isso não quer dizer que os enunciados (2) e (3) não sejam descrições de estados mentais do indivíduo A, embora ela seja uma descrição bem distinta da descrição de objetos visíveis tais como um celular. A primeira questão que surge aqui é:

a) Quais seriam os critérios de distinção entre essas duas formas de descrições?

Segundo a interpretação de Hacker (HACKER, 2000, 40) sobre o segundo Wittgenstein, os enunciados com termos mentais feitos na

primeira pessoa do singular no presente são descrições distintas de uma descrição feita por enunciados tais como a sentença (1) devido as seguintes razões:

- (i) Os enunciados acerca dos próprios estados mentais não são baseados na percepção.
- (ii) Não existem condições de observação nem tão pouco um órgão interno através do qual apreendemos os próprios estados mentais.
- (iii) Não existe um critério de identificação dos estados mentais e, por esse motivo, um sujeito não reconhece ou deixa de reconhecer seus estados mentais.
- (iv) Não podemos verificar os enunciados sobre os estados mentais de modo aproximado, mas apenas refletir um pouco mais sobre os mesmos.
- (v) No âmbito dos estados mentais, só podemos falar com sentido acerca de indecisão, mas não de conhecimento ou ignorância.
- (vi) Esses enunciados não têm como fundamento a evidência. Por esse motivo, não faz sentido perguntar como o sujeito A sabe que sente tristeza enquanto ele sente tristeza.

Conforme Hacker (HACKER, 2000, 41), essas razões não demonstram que uma descrição dos estados mentais seja uma atividade impossível, mas que o jogo de linguagem desses enunciados é bem mais sofisticado do que se imagina. Concordamos que, de fato, o jogo dos conceitos mentais é sofisticado. Como já foi dito: É verdade que, na maioria das vezes, as descrições de nossos estados mentais fogem muito do padrão da sentença (1), na qual o sujeito A descreve seu novo celular. No entanto, não acreditamos que as descrições dos estados mentais tais como em (5) seja distinta do jogo de linguagem (1) devido a todos os motivos apresentados por Hacker. Vejamos então em que sentido discordamos dessa posição independentemente se essa explicação seja de fato a interpretação correta do segundo Wittgenstein.

### **3. Análise das razões**

Antes de apresentarmos algumas críticas para essas razões, devemos analisar primeiramente a distinção entre “descrição” e “expressão” que

parece ser defendida pelo segundo Wittgenstein. Imaginemos que o sujeito A afirme espontaneamente ao psiquiatra:

(7) “Eu tenho dor de cabeça agora”

De acordo com a interpretação de Hacker (2000), o enunciado (7) não seria a rigor uma descrição de seu estado mental da dor, mas apenas sua *expressão* natural. Ao fazer esse proferimento, o sujeito A não observou sua mente, verificou que existe algo lá que é “dor” e, por último, descreveu esse estado mental ao psiquiatra. Em sua opinião, relatos de dor são extensões sofisticadas de um jogo primitivo que é o comportamento da dor. Vejamos, por exemplo, como uma criança geralmente aprende o significado de “dor”. Em primeiro lugar, ele se machuca e depois grita:

(8) “Ai”

Diante do enunciado (8), os pais afirmam que ele está com dor ou lhe perguntam se está com dor. Desse modo, a criança aprende que aquilo que ela sentiu é denominado “dor”. Conforme a leitura de Hacker (HACKER, 2000, 43), deveríamos dizer que tanto (7) quanto (8) não são descrições de estados mentais, mas apenas *expressões* ou *manifestações*, posto que esses proferimentos excluem a dúvida e não expressam assim um modo de conhecimento. Isso não significa que os termos “dor de cabeça” e “ai” sejam idênticos à ação de gritar. Ao contrário, essas palavras substituem esse comportamento (WITTGENSTEIN, 1952, § 243). O sujeito A poderia inclusive apenas ter franzido a testa passando a mão nela e diante do seu comportamento o psiquiatra poderia ter corretamente afirmado:

(9) “Ele está com dor de cabeça”

Os enunciados (7) e (9) se referem à mesma situação, isto é, ao fato de alguém estar sentindo dor de cabeça. A possibilidade do enunciado (9) no jogo de linguagem dos conceitos psicológicos aponta assim para o fato de que há um padrão comum de comportamento de dor que possibilita a

comunicação entre sujeitos portadores de linguagem, os quais experienciam estados mentais. Além disso, poder-se-ia argumentar que o termo “eu” no proferimento (7) parece ser redundante e não informativo, visto que um sujeito B (diferente do sujeito A) não pode, na prática, experienciar a dor de cabeça do sujeito A. O termo “eu” não pode ser substituído por uma descrição de um corpo e pode ser eliminado, pois através da introspecção não conseguimos encontrar um “eu”. Sendo assim, em vez de (7), poderíamos também afirmar:

(10) “Existe uma dor de cabeça”

Ao contrário dessa posição, poder-se-ia alegar que o termo “eu” não pode ser eliminado em (7), pois essa palavra diferencia o sujeito que sente a dor de cabeça daquele que não a sente, como por exemplo, o psiquiatra do sujeito A. Além disso, a existência da dor de cabeça parece exigir a existência de seu portador. Desse modo, surge a segunda questão:

(b) Quem são os portadores dos estados mentais?

Para alguns filósofos, como por exemplo, Frege (2003), os portadores dos estados mentais são os indivíduos que os experienciam. No caso citado, o portador da dor de cabeça seria o sujeito A. No entanto, de acordo com a abordagem de Wittgenstein, o portador dos estados psicológicos seria aquele que os manifesta, que nesse caso, também seria o sujeito A. Mas vale ressaltar que, conforme a posição wittgensteiniana, ser o portador de um estado mental  $x$  não é uma propriedade do estado psicológico  $x$  em si.

Nesse contexto, poderíamos nos perguntar se essa análise não estaria tratando apenas de conceitos mentais e deixando os próprios fenômenos psicológicos de lado. De acordo com esse autor, isso seria um engano, pois é assim que o jogo inicia e que podemos desfazer os problemas que surgem no âmbito da filosofia. Contudo, a distinção entre “expressão” e “descrição” não é tão clara o suficiente e nem sempre podemos dissociar expressões de descrições dos estados mentais. Suponhamos que o

psiquiatra peça ao sujeito A que descreva sua atual dor de cabeça e como resposta obtenha:

(11) “A minha dor de cabeça é leve e em forma de pressão na testa”

Diante do enunciado (11), o psiquiatra acredita que ele tem cefaleia tensional. Nesse caso, o sujeito A busca descrever tanto a intensidade de sua dor quanto a localização da sensação de dor de cabeça. Nesse caso, poderíamos argumentar que esse proferimento seria uma descrição mais próxima do enunciado (1), posto que ele parece ser informativo. Contudo, a declaração (11) não parece ser corrigível como o enunciado (1). Nesse caso, temos assim uma assimetria epistêmica no enunciado (11) que não necessariamente há no proferimento (1).

Mesmo assim, conforme Hacker (2000), faria mais sentido afirmar que os outros conhecem nossos estados mentais melhor do que nós mesmos, posto que estar consciente de uma dor é o mesmo que sentir essa dor. Suponhamos ainda que o indivíduo A afirme:

(12) “Eu sei que tenho dor de cabeça agora”

De acordo com essa abordagem, o enunciado (12) seria análogo ao proferimento (7). Desse modo, Wittgenstein seria contra a tese de que temos um acesso epistêmico privilegiado aos nossos próprios estados mentais e estaria tentando desmistificar a imagem clássica cartesiana da mente de que há uma dicotomia entre “interno” versus “externo”, posto que conheceríamos o mundo interno melhor do que o mundo externo. Nas palavras de Hacker:

Essa imagem clássica se completa quando acrescentamos a concepção errônea de um suposto acesso privilegiado aos nossos estados e processos internos a postulação de que só podemos saber indiretamente como as coisas se passam com os outros – a postulação de que o interno está oculto por trás do “externo”, vale dizer, por trás daquilo que veríamos como meras externalidades comportamentais – os movimentos corporais e os sons da fala. Wittgenstein procurou mostrar que esta também é uma concepção errônea, mas não

porque o interno seja uma ficção, como querem os behavioristas. Muito pelo contrário, o prazer e a dor, a alegria e a tristeza não são apenas comportamentos (HACKER, 2000, 45).

Em nossa opinião, essa ideia é, no mínimo, estranha, pois ele não deixa suficientemente claro em que sentido o sujeito não pode ter um acesso privilegiado aos próprios estados mentais. Por esse motivo, iremos analisar a seguir o conhecimento acerca dos próprios estados mentais. Mas, antes disso, façamos um resumo breve do método proposto por Wittgenstein no âmbito da mente. De fato, essa metodologia apresenta várias vantagens, mas também tem certos limites. Até agora, a análise conceitual dos enunciados citados nos mostra que os motivos (i) e (ii) apresentados por Hacker são corretos. Os proferimentos sobre os estados mentais atuais não têm como base nem a percepção visual nem a introspecção ou observação interna. De fato, não faz sentido afirmar que há um órgão interno, como por exemplo, um terceiro olho, dentro de nosso cérebro que observa e apreende nossos estados mentais. Porém, isso não nos mostra que a terceira razão (iii) seja correta, visto que de algum modo, os nossos estados mentais atuais recaem sobre nossa consciência e podem ser identificados ou reconhecidos pelo portador desses estados psicológicos.

Acreditamos que um dos erros dessa análise é que não podemos abordar os fenômenos mentais de forma isolada, mas de modo holístico. Observem que não sentimos um determinado estado mental *x* de modo isoladamente, pois todo estado psicológico faz parte de uma *rede* de outros estados mentais *y*, *z* etc. Por exemplo: Um sujeito *A* sente tristeza, inquietação e dor de cabeça. Embora, ele não possa observar esses estados mentais atuais por meio de uma introspecção, ele tem acesso a esses estados mentais de modo consciente e não confunde, por exemplo, sua dor de cabeça com sua tristeza. Por esse motivo, ele tanto pode expressá-lo quanto descrevê-lo distinguindo inclusive os diferentes tipos de estados mentais e os objetos intencionais, aos quais esses estados mentais se

referem. Isso nos mostra que deve haver sim algum critério de identificação ou reconhecimento de nossos estados mentais atuais, porém esse critério não é subjetivo, mas intersubjetivo.

Anteriormente a Wittgenstein, Franz Brentano (1874) também criticou a ideia de que temos um acesso aos nossos estados mentais atuais por meio da introspecção. Imaginemos que o indivíduo A tente observar sua tristeza ou inquietação no momento que sente esses estados mentais. Nesse caso, esses sentimentos seriam alterados pelo próprio sujeito e poderiam inclusive deixar de existir. Nesse contexto, a terceira questão que surge aqui é:

(c) Mas como alguém pode identificar ou reconhecer que tipo de fenômeno mental está sendo experienciado, visto que tais fenômenos não são observáveis por meio de sua introspecção?

Em nossa opinião, a resposta para a questão (c) é que enunciados com termos mentais na primeira pessoa do singular envolvem *autoevidência*. Por esse motivo, a dúvida não faz parte do jogo de linguagem dos conceitos mentais. Porém, isso não significa que a ignorância no âmbito da mente seja sempre impossível. Sendo assim, também discordamos das razões (iv), (v) e (vi) apontadas por Hacker (HACKER, 2000, 41).

É um erro achar que questões acerca de como o sujeito A sabe que está triste ou que ele está com dor não fazem parte do jogo de linguagem dos conceitos mentais. Vale lembrar que ao afirmar que estados mentais não podem ser observados por meio da introspecção, Wittgenstein não defende que a introspecção seja sempre impossível no âmbito da consciência. A introspecção faz sentido enquanto uma avaliação dos próprios estados mentais passados (WITTGENSTEIN, 1952, §587). Imaginemos que o sujeito A tenha refletido bastante sobre seus estados mentais antes de procurar o psiquiatra B. Nesse caso, o indivíduo A fez uma introspecção e descreveu seus estados mentais a B. Isso quer dizer

que no caso de estados mentais não atuais, podemos falar de uma descrição de nossos estados mentais sem problemas.

Nesse contexto, Wittgenstein parece estar de acordo com a ideia defendida por Franz Brentano em sua obra *Psychologie vom empirischen Standpunkt* (1874) de que a introspecção dos fenômenos mentais atuais é impossível, visto que ela altera o caráter desses atos mentais, mas que podemos falar com sentido de uma introspecção nos casos de estados mentais não atuais. No entanto, como já argumentamos anteriormente: Apesar de não podermos fazer uma introspecção ou observação interna dos nossos estados mentais atuais, sabemos, a partir de nossa experiência, que esses estados mentais são conscientes e que nós, geralmente, não confundimos um estado mental x com outros estados mentais atuais y, z etc. Além disso, aprendemos não apenas os significados dos conceitos mentais, mas também as referências dos mesmos. Analisemos então de modo mais cuidadoso ainda se, de fato, as razões (iii), (iv), (v) e (vi) apresentadas por Hacker fazem mesmo sentido. Nesse contexto, podemos levantar aqui uma quarta questão:

- (d) Se não houver um critério de identificação dos estados mentais, como pode o sujeito saber que o fenômeno mental que ele sente é, por exemplo, “inquietação”, “dor” ou “tristeza”?

Nesse caso, Wittgenstein diria que palavras tais como “inquietação” ou “tristeza” são aprendidas do mesmo modo que vários outros termos de nossa linguagem, ou seja, por meio do uso dessas palavras em diferentes contextos. O critério para saber se o uso dessas palavras é correto ou não é público e não privado. O problema é que estamos acostumados com o modelo “linguagem-objeto”, ou seja, com a ideia de que a cada palavra deve corresponder um objeto e, por isso, acreditamos erroneamente que a linguagem funciona de uma única forma. Como já foi dito: No âmbito do jogo de linguagem dos conceitos mentais, não há um objeto visível que o sujeito observe e o identifique por meio da introspecção como sendo o

fenômeno x ou y. Por esses motivos, o que é possível observar é apenas o uso que fazemos desses termos.

Discordamos dessa última ideia, isto é, que aquilo que podemos observar é apenas o uso que fazemos dos conceitos mentais. Mas concordamos que a linguagem não funciona de um único modo e que nossos estados mentais não são análogos a objetos comuns como, por exemplo, celulares, cadeiras, mesas, os quais podem ser observados por meio da percepção visual. Não se trata aqui, portanto, de uma Linguagem-Objeto tal como no jogo de linguagem primitivo citado por Wittgenstein (Wittgenstein, 1952, §2). Porém, conceitos mentais tais como “dor” ou “tristeza” devem ter uma referência, embora essa referência não seja por nós ainda determinada. Para Wittgenstein, a referência dos estados psicológicos parece ser os comportamentos de dor e não os processos cerebrais que os causam, posto que, por meio de uma introspecção, não temos como achar tais eventos e reconhecê-los como tal. Se tentarmos entender o jogo de linguagem dos conceitos mentais através do modelo “Linguagem-Objeto”, o objeto se torna irrelevante (WITTGENSTEIN, 1952, §293).

Em nossa opinião, deveríamos utilizar aqui não apenas a análise conceitual dos termos mentais como foi proposta por Wittgenstein, mas também a análise fenomenológica descritiva introduzida por Franz Brentano. As vantagens dessa última abordagem acerca dos estados mentais são: em primeiro lugar, não é necessário definir a natureza dos estados mentais, posto que processos cerebrais não são objetos de estudo da filosofia da mente, mas sim das ciências naturais e cognitivas. Com isso, não pretendemos dizer que o conhecimento científico acerca da mente não é fundamental à filosofia da mente, posto que os conhecimentos dessas duas áreas se complementam. Em segundo lugar, é possível delimitar o objeto de estudo da filosofia da mente sem excluir o fato de que nem todo estado mental possui um conteúdo proposicional. De acordo com abordagem fenomenológica, a linguagem não é a imagem perfeita de nossos estados mentais. Isso significa que, por meio da análise conceitual,

deixamos alguns aspectos essenciais de fora, pois alguns estados mentais são anteriores à nossa aquisição da linguagem. Uma criança sente fome e sede antes mesmo de falar e expressar seus estados mentais por meio do choro. Vamos supor que uma criança esteja com fome. Se a mãe lhe oferecer um brinquedo em vez do leite materno, ela não ficará satisfeita. Estados mentais possuem um conteúdo (não necessariamente proposicional) com condições de satisfação e aspectos subjetivos. Por exemplo: Ao sentir fome, a criança também pode desejar o leite materno como condição de satisfação do seu desejo de saciar a fome. Sendo assim, a mãe também poderia lhe oferecer um suco e a criança poderia continuar insatisfeita e chorar. Nesse caso, o conteúdo de seu desejo seria satisfazer sua fome e o aspecto subjetivo seria que sua fome seja satisfeita com o leite materno.

Como já foi dito: a análise conceitual proposta por Wittgenstein não demonstra que não há critérios para identificarmos nossos estados mentais e os significados de termos mentais, mas que tais critérios não são subjetivos e internos. Em outras palavras: O critério é intersubjetivo. Imaginemos ainda que o sujeito A afirme ao psiquiatra:

(13) “Essa dor de cabeça que estou sentindo agora é mesma dor de cabeça que eu senti ontem”.

Nesse contexto, poderíamos primeiramente fazer a distinção entre (i) *identidade estrita* ou *numérica* e (ii) *identidade lata* ou *qualitativa*. A identidade estrita ou numérica (i) ocorre quando há uma relação de um objeto consigo mesmo, isto é, quando há uma igualdade. Um exemplo da identidade numérica seria “ $2 + 3$  é igual a 5”. Já a identidade lata ou qualitativa (ii) ocorre quando há uma relação de semelhança entre objetos ontologicamente distintos. No caso do enunciado (13), poder-se-ia afirmar que há uma identidade lata ou qualitativa (ii), visto que a expressão “é a mesma” parece significar “é semelhante”. Os critérios de distinção das dores entre si seriam suas características fenomênicas, intensidade e

localização. Nesse sentido, a identidade lata parece fazer sentido no jogo de linguagem dos conceitos mentais.

Já a identidade numérica no âmbito da mente enfrenta certas dificuldades, visto que não podemos apontar para o estado mental em questão e afirmar “é essa dor aqui”. (WITTGENSTEIN, 1952, §253). Além disso, o possuidor seria a característica essencial da própria dor. Contudo, “sentir uma dor” é uma propriedade do portador e não o contrário. Desse modo, no enunciado (13), “sentir uma dor” seria a característica do indivíduo A.

De acordo com a interpretação de Hacker (HACKER, 2000, 27-28), a distinção entre (i) identidade estrita ou numérica e (ii) identidade lata ou qualitativa não se aplica ao jogo de linguagem dos conceitos mentais nem no jogo de linguagem das qualidades, como, por exemplo, das cores. Essa distinção faz sentido apenas no jogo de linguagem de objetos físicos tais como celulares, livros, mesas etc. Isso não significa que o enunciado (13) não faça sentido, mas que a expressão “é a mesma” faz uma identificação entre a dor passada e a dor atual do indivíduo A. Além disso, do ponto de vista lógico, poderíamos afirmar sem problemas que um outro indivíduo B sente a mesma dor que o indivíduo A. Imaginemos que após ouvir a declaração (13), o psiquiatra afirme:

(14) “Ontem eu também senti essa mesma dor de cabeça”.

O enunciado (14) faz sentido nesse contexto, mas a identidade entre as dores do indivíduo A e do psiquiatra não seria nem numérica nem qualitativa e os critérios para tal identificação seriam as propriedades fenomênicas das dores, intensidade e localização. Por esses motivos, Wittgenstein pretende criticar apenas a tese de que os critérios de identificação dos próprios estados mentais sejam subjetivos e internos.

Em nossa opinião, a análise de Wittgenstein dos conceitos mentais não se opõe à ideia de que os estados mentais são experienciados de modo imediato pelo sujeito. A rigor, através dessa investigação, Wittgenstein pretende rejeitar três ideias amplamente difundidas no âmbito da filosofia

da mente. A primeira ideia (i) é a de que podemos observar nossos próprios estados mentais através de uma *introspecção*, a qual também é denominada “observação interna”. A segunda ideia (ii) é a da *posse privada* de nossos estados mentais, ou seja, que o sujeito que experiencia esses estados mentais seja o “possuidor” dos mesmos. A terceira ideia (iii) é a da *privacidade epistêmica*, isto é, a tese de que apenas o sujeito A pode conhecer os seus estados mentais. Os outros indivíduos teriam apenas uma vaga ideia ou podem apenas imaginar o que o sujeito A sente.

No entanto, às vezes, Wittgenstein parece criticar uma quarta tese acerca de nossas mentes: (iv) a autoridade epistêmica. Essa tese significa que aquele que é geralmente considerado como sendo o “possuidor” dos estados mentais deve conhecer melhor os seus estados mentais do que outro indivíduo. Retomemos o exemplo do sujeito A que sente tristeza constante, inquietação e acha que está depressivo e, que por isso, procura um psiquiatra. Se a abordagem de Hacker estiver correta, poderíamos afirmar que, nesse caso, o indivíduo A não pode observar esses estados mentais por meio da introspecção no momento em que ele experiencia esses sentimentos. Contudo, ele pode fazer uma reflexão acerca desses estados mentais. Depois, ele expressa esses estados psicológicos para o psiquiatra por meio da linguagem. Após os seus relatos, o psiquiatra pode avaliar seus enunciados e comportamentos e afirmar, por exemplo, que ele tem ansiedade. Mas vale observar que, nesse sentido, ansiedade não é exatamente um estado mental, mas sim uma doença. Por esse motivo, é possível que o psiquiatra afirme com mais autoridade epistêmica acerca da doença psicológica que A possui do que o próprio sujeito A.

Nesse âmbito, devemos distinguir três teses epistêmicas que geralmente são denominadas “cartesianas”, a saber: (i) tese da *infalibilidade*, (ii) tese da *transparência* e (iii) tese da *incorrigibilidade*. Conforme a tese da infalibilidade (i), se o sujeito A acredita que sente um estado mental x, então ele realmente sente x e sua crença sobre x seria justificada, pois o sujeito A tem um acesso imediato aos seus estados mentais e não pode estar errado acerca dessa crença. De acordo com a tese

da transparência (ii), se o sujeito A sente um estado mental x, então ele também deve ter a crença de que sente x. Isso significa que os estados mentais são transparentes para o próprio “possuidor”. A tese da incorrigibilidade (iii) afirma que a crença do sujeito A de que ele sente x não pode ser nem posta em dúvida nem corrigida por ele mesmo ou outro sujeito. Iremos argumentar aqui que as teses (i) e (iii) são corretas, mas que a tese (ii) é falsa. Por esse motivo, faz sentido falarmos de um conhecimento dos próprios estados mentais. Imaginemos então que o psiquiatra faça o seguinte proferimento:

(15) “Eu sei que ele (o sujeito A) está com dor de cabeça”.

O verbo “saber” no enunciado (15) pode significar “eu não tenho certeza” ou ainda “eu posso duvidar”. Nesse caso, a expressão “eu sei” parece ser utilizada no sentido epistêmico habitual e não uma mera ênfase da própria sensação de dor de cabeça. A questão sobre como ele sabe disso, pode ser levantada aqui. Por esse motivo, faz mais sentido afirmar que o psiquiatra sabe que A sente dor de cabeça, isto é, fazer o enunciado (15), do que afirmar que A sabe que sente dor de cabeça tal como no proferimento (12). Além disso, no jogo dos conceitos mentais, não podemos falar corretamente de um conhecimento direto ou indireto acerca dos próprios estados mentais. Nas palavras de Hacker:

Testemunhar o sofrimento de uma outra pessoa não é ter um conhecimento indireto, e sofrer a dor não é ter um conhecimento direto. Aquilo que o sofredor tem é dor e não conhecimento. Se alguém abre o seu coração para um amigo, este último não poderia dizer, “eu tenho apenas conhecimento indireto de seus pensamentos e sentimentos”. Isto só poderia ser dito caso ele soubesse dos pensamentos do amigo de segunda mão, por ouvir dizer (HACKER, 2000, 47).

Já conforme a interpretação de Michel e Newen (2007, p. 16), acerca do segundo Wittgenstein, em vez de termos uma assimetria de conhecimentos no âmbito da mente, nós teríamos uma assimetria de certezas, pois os outros

podem duvidar se eu estou sentindo dor de cabeça agora, mas eu mesma não. Contudo, tanto eu quanto uma outra pessoa podemos saber que eu estou sentindo dor de cabeça. Nesse sentido, estamos de acordo com essa abordagem, mas não pretendemos decidir aqui qual das interpretações acerca do segundo Wittgenstein é a correta, mas sim analisar a questão independentemente de qual seja a visão de Wittgenstein a respeito disso. A questão que se levanta aqui é:

- (e) Mas poderia o psiquiatra conhecer os estados mentais do sujeito A melhor do que o próprio A?

A resposta à questão acima é: sim, depende do sentido de “conhecer” e da situação. No caso de crianças ou pessoas com distúrbios mentais, podemos falar que nós podemos saber mais acerca de seus estados mentais do que eles mesmos. Mas, analisemos o que os verbos “saber” ou “conhecer” podem significar. Existem diferentes formas de conhecimento. Vale ressaltar que, às vezes, usamos os termos “conhecer” como “saber indubitável” e, às vezes, não. Geralmente, na tradição filosófica, afirma-se que um sujeito A sabe que p quando:

- (1) Ele possui a crença de que p;
- (2) P é verdadeira;
- (3) O conhecimento de A é justificado.

De acordo com essa definição, o conhecimento é uma crença verdadeira justificada. Mas essa não é a única definição desse termo. Porém, como a introspecção dos estados mentais atuais é uma atividade impossível, não faria sentido afirmar que há um conhecimento justificado. No entanto, devemos considerar aqui que o autoconhecimento não é impossível e mesmo que a introspecção dos estados mentais atuais seja impossível, nós podemos obter informações a respeito dos nossos estados mentais por meio da própria consciência da rede dos estados psicológicos que experenciamos. Na opinião de Hacker (2000), Wittgenstein apenas aceitaria falar de

“conhecimento” aqui caso a dúvida fosse possível no jogo de linguagem dos termos psicológicos. Mas, em nossa opinião, a ignorância é possível aqui, posto que os estados mentais não são necessariamente transparentes para o sujeito que os experiencia. Por esses motivos, mesmo que usássemos o termo “conhecimento” tal como Wittgenstein parece aceitar, nós poderíamos afirmar que o conhecimento dos próprios estados mentais é possível e faz sentido no jogo de linguagem dos conceitos mentais.

Antes mesmo de uma criança aprender uma linguagem, ela aprende como o mundo funciona. Isso significa que existem formas de conhecimento que são anteriores à aquisição da linguagem. Além disso, a análise de Wittgenstein não vai contra a tese de que existem estados mentais inconscientes. Mas para compreendermos melhor nossa argumentação, devemos esclarecer a classificação brentaniana dos estados mentais em:

- (I) Representações mentais;
- (II) Juízos;
- (III) Emoções.

A representação mental (I) é o estado mental responsável pela captação do objeto intencional e que o apresenta para a consciência de modo neutro, ou seja, sem que haja algum posicionamento epistêmico do sujeito acerca do objeto apresentado. Já o juízo (II) é o estado psicológico, no qual o indivíduo se posiciona epistemologicamente diante do objeto intencional. É possível que um sujeito tenha uma representação mental (I) sem experimentar um juízo (II) a respeito do objeto intencional, mas não o contrário. Contudo, existem dois tipos de juízos, a saber: (a) juízos evidentes e (b) juízos não evidentes. Os juízos evidentes (a) são aqueles juízos feitos pela consciência do sujeito acerca dos próprios estados mentais atuais. A evidência desse modo de juízo é aceita por Brentano, posto que há uma identidade entre o ato mental e objeto ao qual ele se refere, isto é, não é possível separarmos o ato mental que experiencia do objeto experienciado. No entanto, isso não significa que todo juízo acerca dos próprios estados mentais seja evidente, pois nos casos de estados psicológicos passados, não temos uma evidência.

Já os juízos não evidentes (b) são aqueles feitos pelo sujeito acerca dos fenômenos físicos tais como, por exemplo, um celular. As emoções (III) são os estados mentais, nos quais o objeto mental é amado ou odiado pelo sujeito.

Uma vez que é possível que o indivíduo experiencie um ato mental, como, por exemplo, uma inveja de outro sujeito, e não ter necessariamente um juízo acerca do mesmo, há a possibilidade do sujeito desconhecer tal sentimento. A possibilidade do desconhecimento não vai contra a tese da infalibilidade (i). Nesse caso, o sujeito teria apenas a representação mental desse estado psicológico. Se aceitarmos esse modo de explicação fenomenológica dos nossos estados mentais, podemos então defender que embora as teses “cartesianas” da *infalibilidade* (i) e da incorrigibilidade (iii) serem corretas, isso não significa que a tese da transparência (ii) do mental seja também correta. Desse modo, podemos falar de um conhecimento acerca dos próprios estados mentais sem problemas. Além disso, apesar dos juízos evidentes serem incorrigíveis, o conhecimento acerca dos próprios estados mentais pode ser complementado por meio da sua introspeção dos estados mentais passados, da análise da linguagem e dos relatos dos estados mentais feitos por outros sujeitos. Afinal de contas, nós não somos biologicamente tão distintos a ponto de não podermos saber que alguém ao se machucar sente dor. Não é preciso sentir a dor de cabeça do sujeito A para sabermos que aquilo que ele sente é “dor”.

Por todos os motivos relatados até aqui, não concordamos que a descrição feita no proferimento (1) seja distinta da descrição realizada no enunciado (5) devido a todas as razões apontadas por Hacker (2000) ao interpretar o segundo Wittgenstein.

#### **4. Considerações finais**

A discussão sobre a descrição dos próprios estados mentais nos aponta assim para algumas distinções entre “descrever” nas frases (1), isto é, na descrição de objetos físicos tais como o celular do sujeito A, e

“descrever” no proferimento (5) que se refere aos estados mentais do indivíduo A. Contudo, ao contrário de Hacker (2000), os critérios de distinção entre esses dois modos de descrição são:

- (i) Os enunciados acerca dos próprios estados mentais não são baseados na percepção visual, já os enunciados sobre objetos físicos sim.
- (ii) Não existem condições de observação nem tão pouco um órgão interno (um terceiro olho) através do qual apreendemos os próprios estados mentais, mas apenas dos objetos físicos.

Porém, em nossa opinião:

- (iii) Tanto no caso dos objetos físicos quanto no caso dos estados mentais, existe um critério de identificação desses fenômenos, mas tal critério não é nem interno nem subjetivo.
- (iv) Nós podemos verificar os enunciados sobre os estados mentais tanto por meio da análise linguística quanto por meio da análise fenomenológica e refletirmos tanto sobre os enunciados com termos mentais quanto sobre os enunciados acerca dos objetos físicos.
- (v) Tanto no âmbito dos objetos físicos quanto nos estados mentais, podemos falar com sentido acerca de conhecimento ou desconhecimento (ignorância).
- (vi) As descrições acerca dos próprios estados mentais atuais têm como fundamento a evidência caso o sujeito tenha um juízo acerca do estado mental atual em questão.

Nesse caso, a dúvida não faz sentido nesse jogo de linguagem e o conhecimento acerca do mesmo é incorrigível, mas tal conhecimento não é completo e pode ser complementado por meio da intersubjetividade. Já na descrição dos objetos físicos tais como um celular, não temos juízos evidentes, pois o órgão que experiencia e o objeto experienciado são distintos e os nossos sentidos podem nos enganar. Por esse motivo, a dúvida faz sempre sentido no jogo de linguagem dos objetos físicos.

## 5. Referências

BRANDL, Johannes. **Die Entwicklung der Autorität der Ersten Person**. De Grüter, n.5, p. 937-962, 2014.

BRENTANO, Franz. **Psychologie vom empirischen Standpunkte (1874)**. Von der Klassifikation der psychischen Phänomene (1982). Band 1. Frankfurt: Ontos Verlag, 2008.

CARVALHO, Joelma M. **Linguagem e Intencionalidade**. 1. ed. Guarapuava: Apolodoro Edições Virtuais, 2019.

FREGE, Gottlob. Der Gedanke. Eine logische Untersuchung. In: **Logische Untersuchungen**. Göttingen: Vandenhoeck & Ruprecht, 2003.

GLOCK, Hans-Johann. **Dicionário Wittgenstein**. Rio de Janeiro: Ed. Jorge Zahar, 1998.

GRAYLING, Anthony C. **Wittgenstein**. São Paulo: Edições Loyola, 2002.

HACKER, Peter. **Wittgenstein**. Sobre a natureza humana. São Paulo: Editora UNESP, 2000.

MICHEL, Christoph; NEWEN, Albert. Wittgenstein über Selbstwissen: Verdienste und Grenzen seiner Sprachspielepistemologie. In: MÜLLER, NEWEN, A. (Org.). **Logik, Begriffe und Prinzipien des Handelns**. Paderborn: Mentis, 2007, p. 01-32.

WITTGENSTEIN, Ludwig. **Fichas (Zettel)**. Lisboa: Edições 70, 1989.

\_\_\_\_\_. **O Livro Azul**. Lisboa: Edições 70, 1992.

\_\_\_\_\_. **Investigações filosóficas**. São Paulo: Nova Cultural, 1996 (Coleção Os Pensadores).

\_\_\_\_\_. **Da Certeza**. Lisboa: Edições 70, 1998.

\_\_\_\_\_. **Tractatus Logico-Philosophicus**. São Paulo: Edusp, 2001.



## Pensamento “crítico” & pseudo-história

*Ricardo Seara Rabenschlag*<sup>1</sup>

Em sua incansável luta contra os inimigos da Ciência, os autoproclamados pensadores “críticos” divulgam todos os dias uma enorme quantidade de mitos sobre a História da Ciência, que são mais prejudiciais à Ciência, que eles pretendem defender, do que às concepções pseudocientíficas que eles se empenham em combater, isto porque a principal causa do irracionalismo pseudocientífico é o desconhecimento da natureza da Ciência e não há como falsificar a História da Ciência sem deformar a compreensão que temos da sua natureza.

No presente capítulo, analisaremos cinco mitos sobre o surgimento da Ciência Moderna que têm sido largamente divulgados, em nome de uma suposta defesa da Ciência. Ao longo do texto, farei uso da expressão “pensadores “críticos””<sup>2</sup>, com o adjunto nominal sempre entre aspas, para me referir a certo grupo de ativistas, mais ou menos homogêneo e, em geral, composto por cientistas em final de carreira, filósofos ativistas,

---

<sup>1</sup> Possui graduação, mestrado e doutorado em Filosofia pela Universidade Federal do Rio Grande do Sul (UFRGS), com período sanduíche junto à University of Virginia (EUA). Possui experiência em gestão de ciência e tecnologia, tendo exercido o cargo de Pró-Reitor de Pesquisa e Pós-Graduação junto à Universidade Estadual do Rio Grande do Sul (UERGS). Atualmente, é professor associado da Universidade Federal de Alagoas (UFAL) e desenvolve pesquisas nas áreas de Filosofia e História da Ciência.

<sup>2</sup> O uso das expressões “Pensamento “crítico”” e “pensador “crítico”” devem ser interpretadas analogamente.

jornalistas “científicos” e “educadores” da ciência, que defendem uma forma extremada de cientificismo ao combaterem o que consideram ser a maior ameaça à civilização ocidental: o irracionalismo anticientífico. Evidentemente, há muitos profissionais que ensinam e divulgam a Ciência de forma honesta e sem um viés ideológico, mas, infelizmente, eles não têm o espaço e a visibilidade que têm os pensadores “críticos”.

### **O mito da criação *ex-nihilo* da ciência moderna**

De acordo com uma versão muito difundida sobre o surgimento da Ciência Moderna, o estopim da revolução foi o surgimento de uma nova estrela nos céus da Europa, no ano de 1604. Neste mesmo ano, Johannes Kepler publicou sua obra *Stela Nuova*, onde ele afirmava, com base na observação da diminuição da magnitude do novo objeto celeste, que o Céu não era feito de matéria incorruptível. Seis anos depois, em 1610, Galileu Galilei publica seu livreto *Siderus Nuncios*, relatando a observação de montanhas na Lua, de manchas no Sol, de fases na iluminação de Vênus e de satélites em torno de Júpiter. Tais observações se somavam as evidências descobertas por Kepler para sustentar a tese de que os corpos celestes não eram essencialmente diferentes dos terrestres, o que contradizia frontalmente a Cosmologia tradicional e abria espaço para a criação de uma nova Física baseada na ideia de que todos os corpos do Universo eram feitos da mesma matéria e obedeciam as mesmas leis. Por meio de experimentos com planos inclinados, Galileu mostraria ainda que a antiga Física aristotélica também estava equivocada quanto à dinâmica dos corpos terrestres, pois, ao contrário do que afirmavam os aristotélicos, a queda dos corpos era acelerada e não dependia da massa. Além disso, inspirado pela Astronomia heliocêntrica de Copérnico, Kepler havia desenvolvido uma Astronomia heliocêntrica capaz de prever o movimento dos planetas, com base em apenas três leis. A primeira lei dessa Astronomia nova afirma que os planetas descrevem uma órbita elíptica em

torno do Sol, algo impensável, não apenas no contexto da Astronomia ptolomaica, mas também da copernicana.

O próprio Kepler relutou muito em abandonar o princípio tradicional da circularidade dos movimentos celestes e só o fez após ter revisado minuciosamente os dados observacionais obtidos por Tycho Brahe. Todas estas descobertas extraordinárias, feitas por Kepler e por Galileu, eram devastadoras para a Física antiga, mas ainda não eram suficientes para destroná-la, já que careciam de uma explicação física. A formulação de uma teoria física capaz de explicar e unificar todas essas descobertas só ocorreria em 1687, ano em que Isaac Newton publica o *Philosophia Naturalis Principia Mathematica*. Nesta que, ainda hoje, é considerada a mais importante obra científica de todos os tempos, Newton conclui a Revolução Científica, deflagrada pelo aparecimento de uma nova estrela nos céus da Europa, ao apresentar uma nova teoria física capaz de explicar, de uma só vez e a partir de um único conjunto de leis universais, tanto a queda acelerada dos corpos terrestres como a órbita elíptica dos planetas em torno do Sol.

É uma bela história e quase tudo que se diz ali é verdadeiro, o único problema é que a Física moderna não surgiu do nada. Na realidade, a Física aristotélica já vinha sendo contestada desde a Antiguidade e as críticas não desapareceram na Idade Média. Após a conquista da Pérsia, em 644 D.C., os muçulmanos não apenas absorveram a Filosofia Natural aristotélica, que havia sido levada para lá por Alexandre, O Grande, como a criticaram e a desenvolveram. Todo esse imenso tesouro intelectual aportou no Ocidente, no ano de 711 D.C., quando os muçulmanos invadiram a península ibérica. A partir desse momento, o mundo cristão tomou conhecimento da Filosofia Natural de Aristóteles, por meio da tradução das obras de Aristóteles do árabe para o latim, bem como dos comentários de Al Farabi, de Avicena e de Averróis, entre outros. Igualmente importantes para a Revolução Científica, foram as críticas de pensadores medievais como Richard Swineshead, Jean Buridan, Nicole Oresme, Domingos de Soto, entre outros. Muito antes do século XVI, já se

suspeitava que a Física de Aristóteles necessitava de ajustes importantes, tanto em relação à explicação do movimento dos corpos terrestres, quanto à dos corpos celestes. As críticas de Ansazel ao geocentrismo ptolomaico e a teoria do *impetus* de Jean Buridan são dois exemplos que mostram, claramente, que a permanência de uma tradição não exclui a existência de críticas<sup>3</sup>. Portanto, não é razoável falarmos de uma Revolução Científica moderna sem mencionarmos que ela foi a etapa final de um processo que teve início no final da Antiguidade e que se estendeu ao longo de toda a Idade Média. Como observou Thomas Kuhn, em sua obra *The Structure of Scientific Revolutions*, o período de crise de um paradigma pode ser bastante longo, pois, independentemente da gravidade das críticas dirigidas contra ele, a comunidade científica só irá abandonar o velho paradigma depois que ela tiver produzido um novo.

### **O mito da cosmologia heliocêntrica de Copérnico**

Feitas estas breves considerações acerca do longo processo de maturação da revolução científica moderna origem, estamos em condições de avaliar melhor o conteúdo da publicação, em 1543, do *De Revolutionibus Orbium Coelestium* de Nicolau Copérnico. Esta breve análise da obra de Copérnico nos dará a oportunidade de desmascarar um dos principais mitos sobre a origem da Ciência moderna: a tese de que a Astronomia heliocêntrica de Copérnico era um tratado de Cosmologia.

Ao mostrar de forma rigorosa que o movimento observável dos corpos celestes era compatível com um modelo astronômico heliocêntrico, Copérnico iniciou aquilo que, seguindo Alexandre Koiré<sup>4</sup>, prefiro chamar de Revolução Astronômica. A simplicidade e a elegância matemáticas do modelo cuja eficácia ele demonstrou de forma extremamente minuciosa atestavam sua superioridade em relação ao modelo geocêntrico defendido

---

<sup>3</sup> A importância da recepção de Aristóteles no mundo antigo, especialmente entre os árabes, pode ser vista na coletânea de artigos editada por Richard Sorabji que consta na bibliografia.

<sup>4</sup> *The Astronomical Revolution*. Revised edition, Harvard University Press, 1992.

pelos ptolomaicos. Afinal, como o objetivo da Astronomia era salvar os fenômenos, o único critério que um astrônomo poderia utilizar para decidir entre dois modelos experimentalmente equivalentes era a sua perfeição matemática<sup>5</sup>.

Em que pese a opinião dos pseudo-historiadores da Ciência, a Astronomia no século XVI ainda era considerada uma arte matemática e a argumentação copernicana era estritamente geométrica.<sup>6</sup> A tradição medieval reconhecia a existência de quatro artes matemáticas: a Aritmética, a Geometria, a Acústica e a Astronomia. Estas quatro disciplinas compunham o *quadrivium* que, juntamente com o *trivium*, composto de Lógica, de Gramática e de Retórica, formavam o *curriculum* das Universidades durante o final da Idade Média e o início da Moderna. Enquanto a Aritmética e a Geometria eram pensadas como Matemática pura, a Acústica e a Astronomia eram pensadas como Matemática aplicada, fundadas tanto na razão como na observação: a Acústica tendo como objeto o som produzido pelos objetos terrestres e a Astronomia, o movimento dos objetos celestes. Note que a Acústica e a Astronomia não eram consideradas disciplinas teóricas e, sim, técnicas, já que não buscavam explicar a causa dos fenômenos naturais, mas apenas criar modelos matemáticos que ordenassem o aparente caos do mundo natural<sup>7</sup>.

Em outras palavras, ao contrário do que sugerem muitos livros de “História” da Ciência, a aplicação da Matemática à natureza não é uma invenção moderna. Copérnico era um matemático e nunca pretendeu

---

<sup>5</sup> Na verdade, o modelo ptolomaico só foi superado, em precisão, pelo modelo de Tycho Brahe, no qual a Lua e o Sol orbitavam a Terra, enquanto que os demais planetas orbitavam o Sol: um modelo híbrido que, segundo o próprio autor, visava combinar à perfeição geométrica do modelo copernicano com a correção filosófica do modelo ptolomaico.

<sup>6</sup> O principal argumento contra esta interpretação é a alegação de que Copérnico temia ser perseguido pela Igreja, razão pela qual teria decidido não publicar sua obra em vida. O problema óbvio com esta interpretação é que a obra postumamente publicada é estritamente astronômica e que os argumentos apresentados eram estritamente matemáticos. Copérnico era um Cônego da Igreja Católica e, de fato, temia uma reação negativa por parte dos inquisidores da Igreja, o que explica o fato de ele não ter publicado sua obra em vida. Mas a razão desse temor não é a de que seu livro fosse um tratado de Cosmologia, mas sim que ele fosse interpretado como tal.

<sup>7</sup> Certamente, esta é a concepção platônica da natureza da Acústica e da Astronomia. Aristóteles não apenas concorda com Platão, nesse aspecto, como também inclui a Mecânica como exemplo de *tecné* matemática.

provar que o Sol se encontrava no centro do Universo, embora ele acreditasse nisso e tivesse consciência de que sua Astronomia heliocêntrica poderia ser usada como parte de uma argumentação a favor de uma Cosmologia heliocêntrica. Uma vez que Martinho Lutero recém havia sido condenado por heresia, ele também tinha receio de que sua obra pudesse ser mal compreendida e decidiu não a publicar em vida. Ironicamente, não foi o Papa e sim os reformistas Lutero e Calvino que condenaram o livro de Copérnico. A condenação pela Igreja Católica só ocorreria no século seguinte, em 1616, motivada pela publicação das obras de Galileu que atribuíam ao astrônomo Copérnico a defesa de uma Cosmologia heliocêntrica, o que contrariava os ensinamentos dos doutores da Igreja. Entretanto, em 1757, o Papa Bento XIV retirou a obra de Copérnico do *Index Librorum Prohibitorum*, uma vez que a controvérsia entre heliocentristas e geocentristas não contrariava nenhum dogma da Igreja Católica e que a divergência entre os doutores já havia sido resolvida por Newton, em favor de uma Cosmologia heliocêntrica.

### **O mito da perda da posição privilegiada da terra**

Sempre que um pensador “crítico” se aventura a falar sobre as origens da Ciência moderna, a Revolução Científica é invariavelmente descrita como um rebaixamento tanto da Terra, que teria sido deslocada da posição privilegiada que ela ocupava no Universo, como do Céu, que teria se relevado igualmente corruptível e constituído da mesma matéria de que somos feitos. Tudo isso é muito comovente, mas pouco sério, revelando apenas a falta de honestidade intelectual destes pseudo-historiadores da Ciência. O viés ideológico é particularmente evidente na afirmação de que o fato de a Terra ter deixado de ocupar uma posição central na “Cosmologia” copernicana representaria um rebaixamento do lugar do homem no Cosmos. Como vimos na seção anterior, a Astronomia copernicana não é uma Filosofia da Natureza e muito menos uma Cosmologia. Mas mesmo que se entenda a referência como sendo dirigida

aos astrônomos modernos, que realmente defendem a tese de que a Terra gira sobre seu próprio eixo e orbita o Sol, o heliocentrismo jamais representaria um rebaixamento do lugar da Terra no Universo. Isto porque, na Cosmologia medieval, a região central era literalmente o pior lugar do Universo, onde se encontravam as substâncias menos nobres, sujeitas à criação e à destruição, porque constituídas de matéria corruptível. Não foi outra a razão que levou Dante a situar o Inferno no centro da Terra, que ele também considerava como sendo o lugar mais baixo e menos perfeito do Universo. Que a posição central é a mais baixa e corrompida, um verdadeiro vale de lágrimas, simbolizado pela cruz, é algo que deveria ser tão óbvio para um ocidental que o fato de muitos estudantes acreditarem nesse mito propagado pelo pensamento “crítico” demonstra quão pobre é a educação que eles receberam, não apenas em relação à História da Ciência, mas sobretudo em relação à História do Ocidente.

Quanto ao suposto rebaixamento dos objetos celestes, que teriam perdido seu estatuto especial, o que ocorreu foi exatamente o oposto, pois, o que antes valia apenas para a matéria celeste, a descrição puramente quantitativa, passou a valer também para a matéria de todos os corpos, fossem eles celestes ou terrestres<sup>8</sup>. A elevação da matéria terrestre à mesma categoria da matéria celeste<sup>9</sup> foi o início do fim da Física aristotélica, pois, ao contrário de um planeta, uma bola de metal pode facilmente ser lançada de um canhão, largada do alto de uma torre ou do topo de um plano inclinado. Para Galileu, o mais ardoroso defensor moderno da Cosmologia “copernicana”, tudo se passa como se todos os corpos fossem feitos de matéria celeste, e todo objeto físico é tratado como *res extensa* e descrito unicamente por suas qualidades primárias. Foi

---

<sup>8</sup> Pode-se alegar que, na Física aristotélica, a explicação do movimento das esferas celestes não era inteiramente mecânica, o que é verdadeiro. Contudo, não estamos afirmando que a Física celeste aristotélica já era uma Física moderna, mas apenas que ela já era uma Física matemática, ainda que não o fosse exclusivamente.

<sup>9</sup> Não estamos afirmando que, para os físicos modernos, os corpos terrestres são substâncias incorruptíveis. Tampouco seria correto afirmar que, para os físicos modernos, os corpos celestes são substâncias corruptíveis. A Revolução Científica moderna representa um abandono da concepção hilemórfica de substância natural, o que é particularmente evidente na Física cartesiana.

justamente a crença de que a tendência natural ao movimento circular que Galileu, seguindo a tradição pitagórico-platônica, atribuía aos corpos celestes valia, igualmente, para os corpos terrestres que o levou a sustentar um princípio geral de inércia circular, com o qual ele acreditava poder explicar a translação e a rotação da Terra, assim como o fenômeno das marés, que ele via como um efeito da rotação, combinada ao formato esférico da Terra, igualmente atribuído a essa misteriosa tendência ao movimento circular<sup>10</sup>.

A revolução científica moderna, sobretudo na sua fase inicial, representa um profundo distanciamento da tradição aristotélica e uma reaproximação da filosofia pitagórica<sup>11</sup>. Contra a antiga física aristotélica, os novos físicos defendiam que a natureza é composta de um único tipo de matéria que se deixa descrever em termos puramente matemáticos. Para usarmos a bela metáfora de Galileu, o lema fundamental da racionalidade científica moderna é de que o livro da natureza está escrito em caracteres matemáticos. Não se trata de afirmar que só existam substâncias naturais ou que a crença no sobrenatural deve ser descartada como irracional, pois nenhum dos precursores da ciência moderna subscreveu semelhante tese. Aliás, a bela metáfora de Galileu surge da comparação que ele faz entre os livros sagrado e científico.

## **O mito do dogmatismo dos doutores da igreja**

Os pensadores “críticos” acusam os doutores da igreja de dogmatismo por eles terem se recusado a olhar pelo telescópio de Galileu, uma atitude que eles também classificam de anticientífica<sup>12</sup>. Estes bravos guerreiros, em geral muito cultos em matérias científicas, demonstram

---

<sup>10</sup> Posteriormente, Galileu parece ter revisado sua posição e passado a defender uma inércia linear, a partir do resultado de um experimento mental feito com base nas suas experiências com o plano inclinado.

<sup>11</sup> Copérnico teve grande influência do pitagorismo, uma corrente filosófica que se opôs ao naturalismo qualitativo dos milesianos e que também influenciou profundamente o platonismo.

<sup>12</sup> Esta alegação se baseia em uma carta que Galileu escreveu a Kepler se queixando de que seus colegas da Universidade sequer se dignavam a olhar pelo seu telescópio.

uma miopia muito acentuada quando o assunto é História e Filosofia da Ciência. Sem entrarmos no mérito da veracidade deste relato “histórico”, não resta dúvida de que é falsa a tese pressuposta na acusação feita contra os doutores da igreja, a saber, que Galileu e os doutores da Igreja concordavam acerca do que constituía um fato científico. Não sendo verdadeira esta tese, a grave acusação perde toda sua força. Com efeito, se apenas o resultado da observação direta da realidade for admitido como evidência científica, critério adotado pelos doutores da Igreja e por todos os filósofos da natureza desde Tales, as observações feitas com o uso do telescópio não podem contar como evidência, o que justificaria a suposta recusa dos doutores da igreja em olhar através do estranho instrumento construído por Galileu.

Na verdade, nem precisamos apelar para a diferença entre as concepções tradicional e moderna de evidência científica para mostrar quão tola é a crítica feita aos doutores da Igreja. Com efeito, a própria metodologia científica moderna só considera como fato as observações feitas com base em experimentos científicos e tais experimentos devem, necessariamente, ser realizados por meio de instrumentos de medida cujo modo de funcionamento e limite de precisão sejam conhecidos previamente. Como na época em que ocorreu o suposto evento envolvendo Galileu e os doutores da Igreja ainda não existia uma teoria óptica capaz de explicar o que acontecia com a luz dentro do estranho aparelho construído por Galileu, o telescópio ainda não se qualificava como um instrumento científico e as observações feitas com ele não deveriam ser aceitas como evidências científicas.

A situação é particularmente embaraçosa, já que muitos astrônomos renomados, incluindo Tycho Brahe, tinham conhecimento da existência da luneta, muito antes de Galileu ter construído uma para tentar vendê-la aos mercadores de Veneza, e todos, incluindo Kepler, eram contrários ao seu uso na Astronomia, justamente, porque as leis ópticas capazes de explicar seu funcionamento ainda não eram conhecidas, o que impossibilitava a determinação do grau de confiabilidade das observações feitas com o uso

de tal instrumento. Somente com a publicação do tratado de óptica de Kepler, em 1611, este requisito foi satisfeito e, finalmente, se pode explicar o funcionamento do telescópio e determinar o grau de precisão das observações feitas por meio dele, o que permitiu a Kepler aprimorar consideravelmente o telescópio refrator. Ainda assim, em razão da enorme dificuldade técnica envolvida na produção de lentes confiáveis, o uso do telescópio na Astronomia só viria a se consolidar após a invenção do telescópio refletor, feita por Newton, por volta de 1680. Ao usar um espelho ao invés de uma lente para captar a luz, os telescópios de Newton evitavam as aberrações cromáticas e outras imperfeições produzidas pelas lentes objetivas usadas nos telescópios refratores, o que permitiu a construção de telescópios muito mais potentes e precisos.

### **O mito da incompatibilidade entre ciência e religião**

O mito de que há uma incompatibilidade essencial entre Ciência e Religião aparece sob duas formas principais: uma na qual se afirma que a Igreja é contra a Ciência e outra, na qual se diz que a fé é contra a razão. No primeiro caso, o alvo principal são as igrejas cristãs, em especial a Igreja Católica<sup>13</sup>; no segundo, o alvo principal são os cristãos, em especial os católicos. Já vimos, no caso dos doutores que, supostamente, se recusaram a olhar pelo telescópio de Galileu, que a acusação nem sempre é procedente. As relações entre fé e razão foram objeto de intenso debate durante toda a Idade Média que, longe de ter sido uma idade de trevas, foi uma época na qual os estudos de Lógica, de Gramática e de Retórica produziram um refinamento e uma sofisticação da arte da análise conceitual e do debate racional que foram cruciais para a superação da Filosofia Natural de Aristóteles.

---

<sup>13</sup> Em regra, a atitude dos pensadores “críticos” em relação a outras Religiões é muito mais branda, especialmente no caso do Budismo e das chamadas religiões da Nova Era.

Nos séculos VII e VIII, a invasão árabe da Península Ibérica resultou na criação das primeiras Universidades e de inúmeras outras instituições cruciais para o avanço da Filosofia, da Ciência e das Artes. Esse iluminismo medieval deixou uma profunda marca no pensamento filosófico com reflexos que se estenderam até o início da Idade Moderna, quando Copérnico redescobriu, por influência da escola pitagórica, a antiga trilha aberta por Aristarco de Samos<sup>14</sup>. Em resposta às dificuldades enfrentadas pelo naturalismo milesiano, a escola pitagórica defendia uma Astronomia puramente matemática que, ao contrário de buscar as causas dos movimentos celestes, contentava-se em relevar sua harmonia, a exemplo do que Pitágoras havia feito com a Acústica. Essa concepção da Astronomia como Matemática aplicada representava um distanciamento da Física aristotélica e uma reaproximação com a Filosofia platônica e sua compreensão mítica da origem e da formação do mundo natural. A disputa dos físicos modernos com os doutores da Igreja, defensores da Física aristotélica, não era apenas uma disputa no campo da Filosofia Natural, tratava-se, sobretudo, de uma disputa acerca da legitimidade da Teologia Natural. Mais do que a compreensão do mundo natural, estava em jogo a compreensão da relação entre o natural e o sobrenatural. A nova Física ameaçava uma vez mais expulsar a natureza do domínio da Filosofia, rebaixando o estudo dos fenômenos naturais a uma arte matemática e reintroduzindo as narrativas míticas como única forma de compreensão da relação entre o natural e o sobrenatural<sup>15</sup>.

Um resultado importante, mas pouco divulgado, da Revolução Científica, é o fato de a Física newtoniana ter restabelecido não apenas a legitimidade da Filosofia Natural, mas também da Teologia Natural, embora não mais sob bases tomistas. Durante muito tempo, a obra teológica de Newton foi deixada em segundo plano ou mesmo

---

<sup>14</sup> No século III a.C., Aristarco de Samos havia proposto uma cosmologia na qual a Terra e o Sol giravam em torno do seu próprio eixo, enquanto orbitavam um fogo central.

<sup>15</sup> Ironicamente, para os pensadores “críticos”, nenhum dos novos físicos defendia uma concepção cientificista. Ainda que, ao longo da História da Ciência moderna, tenha havido muitas controvérsias acerca das relações entre Ciência e Religião, nenhum dos pensadores que contribuíram decisivamente para o desenvolvimento da Física moderna descartou a Religião como uma forma legítima de compreensão da realidade.

desconsiderada, como se não pudesse haver nenhuma conexão relevante entre Ciência e Teologia. A Igreja Católica, que tem sua imagem constantemente denegrada pelos pensadores “críticos”, apesar do papel essencial que as Universidades tiveram na formação da atmosfera intelectual que propiciou o surgimento da Ciência Moderna<sup>16</sup>, não deixou de reconhecer essa conexão e passou a fomentar a nova Ciência Natural dos modernos. Prova disso é o fato de que algumas das principais descobertas científicas dos últimos dois séculos resultaram dessa nobre iniciativa: a formulação da primeira teoria genética, publicada pelo frade Gregor Mendel, num artigo de 1866, no qual ele também anunciava a descoberta dos genes; e a formulação da primeira teoria cosmológica moderna, publicada pelo padre George Lemaître, num artigo de 1927, no qual ele previa não apenas a expansão do Universo, posteriormente comprovada por Edwin Hubble, ganhador do Prêmio Nobel, mas também sua evolução a partir do que Lemaître denominou “átomo primitivo” e que ainda hoje é o núcleo essencial da teoria cosmológica padrão, popularmente chamada de “Teoria do Big Bang”.

Diante destas descobertas magníficas, feitas por clérigos católicos, fica óbvio que a visão difundida pelos pensadores “críticos” de que a Igreja Católica teve um papel eminentemente negativo na Revolução Científica moderna nada mais é do que propaganda ideológica travestida de História da Ciência. Afirmar que o pensamento religioso é inimigo do pensamento científico ou que a crença no sobrenatural é irracional, como afirmam os cientificistas extremos, é fechar os olhos para o fato de que as Igrejas cristãs, especialmente a Igreja Católica, contribuíram de maneira decisiva para a formulação da moderna visão científica do homem e do seu lugar no Universo.

---

<sup>16</sup> Entre os inúmeros estudos minuciosos e fartamente documentados sobre a influência do pensamento antigo e medieval na formulação da visão científica moderna podem ser encontrados nas obras de Stephen Gaukroger, Edward Grant e James Hannam, listadas nas referências.

## Observações finais

Como pretendemos demonstrar por meio desta breve análise de alguns dos principais mitos sobre a História da Ciência moderna, divulgados por pseudo-historiadores que não se cansam de exibir sua erudição em matérias que eles acreditam conhecer mesmo sem ter lido um único livro sério sobre o assunto, o passado da Ciência está sendo falsificado em grande escala por um número crescente de ativistas que se dizem defensores da Ciência. Esses autoproclamados pensadores “críticos” parecem desconhecer o óbvio: quanto pior for nossa compreensão do passado da Ciência, pior será nossa compreensão do seu estado presente, o que, certamente, terá consequências desastrosas para o nosso futuro, já que vivemos numa sociedade tecnocientífica.

Falsificar a História é uma estratégia política bastante antiga e extremamente eficaz, sobretudo em regimes autoritários. Todo aquele que valoriza a Ciência deve valorizar também a liberdade de pensamento, traço essencial das sociedades democráticas, pois, na falta dessa liberdade, a busca da verdade, traço essencial da investigação científica, transforma-se numa mera busca da unanimidade, traço essencial dos regimes autoritários. Por tudo que estamos vendo na atualidade, em relação à sistemática falsificação da História da Ciência, convém que não nos esqueçamos da advertência feita por George Orwell, no prefácio de seu romance *Nineteen Eighty-Four*: “Quem controla o passado, controla o futuro. Quem controla o presente, controla o passado”.

## Referências

BURIDAN, Jean. **Quaestiones super libris quattuor de caelo et mundo**. Boston: Medieval Academy of America, 2012.

COPERNICO, Nicholas. **On the revolutions of the heavenly spheres**. Amherst: Prometheus Books, 1995.

DANTE, Alighieri. **Divina comedia – inferno**. Milano: Mondatori, 2016.

FEYERABEND, Paul. **Against method**. London: Humanities Press, 1975.

GALILEI, Galileo. **Dialogues concerning two new sciences**. New York: Dover, 1954.

\_\_\_\_\_. **Sidereus nuncius**. Venezia: Marsilio Editori, 1997.

\_\_\_\_\_. **Dialogo sopra i due massimi sistemi del mondo**. Milano: BUR Rizzoli, 2014.

GAUKROGER, Stephen. **The emergence of a scientific culture: Science and the Shaping of Modernity 1210-1685**. Oxford: Clarendon Press, 2006.

GRANT, Edward. **A History of natural philosophy**. Cambridge: Cambridge University Press, 2007.

HANNAM, James. **The genesis of science**. Washington: Regnery Publishing, 2011.

KEPLER, Johannes. **Astronomia nova**. Revised Edition. Santa Fe: Green Lion Press: 2015.

KOYRÉ, Alexandre. **The astronomical revolution**. New York: Dover, 1992.

KUHN, Thomas. **The copernican revolution**. Cambridge: Harvard University Press, 1957.

\_\_\_\_\_. **The structure of scientific revolutions**. Second Edition. Chicago: The University of Chicago Press, 1970.

NEWTON, Isaac. **Mathematical principles of natural philosophy and the system of the world**. Ed. Florian Cajori. Berkeley: University of California Press, 1934.

ORWELL, George. **Nineteen eighty-four**. London: Everyman's Library, 1992.

SORABJI, Richard (Ed.). **Aristotle transformed: the ancient commentators and their influence**. Ithaca: Cornell University Press, 1990.

## Sobre os Organizadores



Natural de Brasília, Marcus José Alves de Souza é Doutor em Filosofia pela Universidade Federal da Paraíba, Professor do Instituto de Ciências Humanas, Comunicação e Artes e do Mestrado em Filosofia da Universidade Federal de Alagoas. Líder do Grupo de Pesquisa Linguagem e Cognição.



Potiguar de Mossoró, Maxwell Morais de Lima Filho é Doutor em Filosofia pela Universidade Federal do Ceará e Professor do Instituto de Ciências Humanas, Comunicação e Artes da Universidade Federal de Alagoas desde 2011.

A Editora Fi é especializada na editoração, publicação e divulgação de pesquisa acadêmica/científica das humanidades, sob acesso aberto, produzida em parceria das mais diversas instituições de ensino superior no Brasil. Conheça nosso catálogo e siga as páginas oficiais nas principais redes sociais para acompanhar novos lançamentos e eventos.



**[www.editorafi.org](http://www.editorafi.org)**  
**[contato@editorafi.org](mailto:contato@editorafi.org)**