



Editora
UFPel

Série Investigação Filosófica
Textos Seleccionados de
**Metafísica,
Linguagem e Mente**

Rodrigo Reis Lastra Cid
Pedro Merluzzi
(Organizadores)

DISSERTATIO
FILOSOFIA

**TEXTOS SELECIONADOS DE
METAFÍSICA, LINGUAGEM E MENTE**

SÉRIE INVESTIGAÇÃO FILOSÓFICA

**TEXTOS SELECIONADOS DE
METAFÍSICA, LINGUAGEM E MENTE**

Rodrigo Reis Lastra Cid
Pedro Merluzzi
(Organizadores)



Pelotas, 2020

REITORIA

Reitor: Pedro Rodrigues Curi Hallal

Vice-Reitor: Luís Isaias Centeno do Amaral

Chefe de Gabinete: Taís Ullrich Fonseca

Pró-Reitor de Graduação: Maria de Fátima Cóssio

Pró-Reitor de Pesquisa e Pós-Graduação: Flávio Fernando Demarco

Pró-Reitor de Extensão e Cultura: Francisca Ferreira Michelin

Pró-Reitor de Planejamento e Desenvolvimento: Otávio Martins Peres

Pró-Reitor Administrativo: Ricardo Hartlebem Peter

Pró-Reitor de Infraestrutura: Julio Carlos Balzano de Mattos

Pró-Reitor de Assuntos Estudantis: Mário Renato de Azevedo Jr.

Pró-Reitor de Gestão Pessoas: Sérgio Batista Christino

CONSELHO EDITORIAL DA EDITORA DA UFPEL

Presidente do Conselho Editorial: João Luis Pereira Ourique

Representantes das Ciências Agrônômicas: Guilherme Albuquerque de Oliveira Cavalcanti

Representantes da Área das Ciências Exatas e da Terra: Adelir José Strieder

Representantes da Área das Ciências Biológicas: Marla Piumbini Rocha

Representante da Área das Engenharias e Computação: Darci Alberto Gatto

Representantes da Área das Ciências da Saúde: Claiton Leoneti Lencina

Representante da Área das Ciências Sociais Aplicadas: Célia Helena Castro Gonsales

Representante da Área das Ciências Humanas: Charles Pereira Pennaforte

Representantes da Área das Linguagens e Artes: Josias Pereira da Silva

EDITORA DA UFPEL

Chefia: João Luis Pereira Ourique (Editor-chefe)

Seção de Pré-produção: Isabel Cochrane (Administrativo)

Seção de Produção: Gustavo Andrade (Administrativo)

Anelise Heidrich (Revisão)

Ingrid Fabiola Gonçalves (Diagramação)

Seção de Pós-produção: Madelon Schimmelpfennig Lopes (Administrativo)

Morgana Riva (Assessoria)



GRUPO DE PESQUISA INVESTIGAÇÃO FILOSÓFICA (UNIFAP/CNPq)

O Grupo de Pesquisa Investigação Filosófica (DPG/CNPq) foi constituído por pesquisadores que se interessam pela investigação filosófica nas mais diversas áreas de interesse filosófico. O grupo foi fundado em 2010, como grupo independente, e se oficializou como grupo de pesquisa da Universidade Federal do Amapá em 2019.

MEMBROS PERMANENTES DO GRUPO

Aluizio de Araújo Couto Júnior
Bruno Aislã Gonçalves dos Santos
Cesar Augusto Mathias de Alencar
Daniel Schiochett
Daniela Moura Soares
Everton Miguel Puhl Maciel
Guilherme da Costa Assunção Cecilio
Kherian Galvão Cesar Gracher
Luiz Helvécio Marques Segundo
Luiz Maurício Bentim da Rocha Menezes
Mayra Moreira da Costa
Paulo Roberto Moraes de Mendonça
Pedro Merlussi
Rafael César Pitt
Renata Ramos da Silva
Rodrigo Alexandre de Figueiredo
Rodrigo Reis Lastra Cid
Sagid Salles
Tiago Luís Teixeira de Oliveira

Mais informações no site: <http://dgp.cnpq.br/dgp/espelhogrupo/513418>

Série Investigação Filosófica

A Série Investigação Filosófica, uma iniciativa do **Núcleo de Ensino e Pesquisa em Filosofia** do Departamento de Filosofia da UFPel e do **Grupo de Pesquisa Investigação Filosófica** do Departamento de Filosofia da UNIFAP, sob o selo editorial do NEPFil online e da Editora da Universidade Federal de Pelotas, tem por objetivo precípuo a publicação da tradução para a língua portuguesa de textos selecionados a partir de diversas plataformas internacionalmente reconhecidas, tal como a *Stanford Encyclopedia of Philosophy* (<https://plato.stanford.edu/>), por exemplo. O objetivo geral da série é disponibilizar materiais bibliográficos relevantes tanto para a utilização enquanto material didático quanto para a própria investigação filosófica.

EDITORES DA SÉRIE

Rodrigo Reis Lastra Cid (GP-IF/UNIFAP)
Juliano Santos do Carmo (NEPFIL/UFPEL)

COMISSÃO TÉCNICA

Juliano Santos do Carmo (Diagramador/Capista)
Danilo Jose Ribeiro de Oliveira (Diagramador)
Bruno Borges Moura (Diagramador)

ORGANIZADORES DO VOLUME

Rodrigo Reis Lastra Cid (UNIFAP)
Pedro Merlussi (UNICAMP)

TRADUTORES

Bruno Borges Moura (UNIFAP)
Lucas Roisenberg Rodrigues (PUCRS)
Luiz Helvécio Marques Segundo (ABFR)
Marcelo Fischborn (IFF)
Marco Aurélio Alves (POSDEFIL / UFOP)

REVISORES

André Abath (UFMG)
Jerzy André Brzozowski (UFSC)
Pedro Merlussi (FAPESP)
Sérgio R. N. Miranda (UFOP)

Acesse o acervo completo em: <https://wp.ufpel.edu.br/nepfil/serie-investigacao-filosofica/>

© Série Investigação Filosófica, 2020

Universidade Federal de Pelotas
Departamento de Filosofia
Núcleo de Ensino e Pesquisa em Filosofia
Editora da Universidade Federal de Pelotas

Universidade Federal do Amapá
Departamento de Filosofia
Grupo de Pesquisa Investigação Filosófica

NEPFI online

Rua Alberto Rosa, 154 – CEP 96010-770 – Pelotas/RS

Os direitos autorais estão de acordo com a Política Editorial do NEPFI online. As revisões ortográficas e gramaticais foram realizadas pelos tradutores e revisores. A autorização para a tradução dos verbetes da *Stanford Encyclopedia of Philosophy* neste volume foi obtida pelo *Grupo de Pesquisa Investigação Filosófica*.

Primeira publicação em 2020 por NEPFI online e Editora da UFPel.

Dados Internacionais de Catalogação

N123 Textos selecionados de metafísica, linguagem e mente.
[recurso eletrônico] Organizadores: Rodrigo Reis Lastra Cid; Pedro Merluzzi – Pelotas:
NEPFI Online, 2020.
192p. - (Série Investigação Filosófica).
Modo de acesso: Internet
<wp.ufpel.edu.br/nepfi>
ISBN: 978-65-86440-12-6

1. Metafísica. 2. Linguagem 3. Mente I. Cid, Rodrigo Reis Lastra. II. Merluzzi, Pedro.

COD 100



Para maiores informações, por favor visite nosso site wp.ufpel.edu.br/nepfi

SUMÁRIO

Sobre a Série Investigação Filosófica	10
Introdução	11
(I) Condições Necessárias e Suficientes	14
Robin Smith	
1. Filosofia e Condições	15
2. A Teoria Padrão: Funções de Verdade e Reciprocidade	16
3. Problemas para a Teoria Padrão	18
4. Inferências, Razões Para Pensar, e Razões Pelas Quais	21
5. Conclusão	29
Referências	30
(II) Referência	32
Eliot Michaelson & Marga Reimer	
1. Introdução	33
2. Nomes Próprios	35
2.1. Teorias Descritivistas	36
2.2. Herdeiros Millianos	41
2.3. Nomes com Múltiplos Portadores	43
3. Indexicais	46
3.1. Indexicais Puros	47
3.2. Indexicais Impuros	51
4. Descrições Definidas	58
5. Modelos da Referência Linguística	64
6. Visões Negativas da Referência	67
Referências	70
(III) A Metafísica da Causação	75
Jonathan Schaffer	
1. Relata	76

1.1. Imanência	79
1.2. Individuação	82
1.3. Aridade	87
1.3.1. Contrastividade	87
1.3.2. Relata Secundário	91
2. Relação	93
2.1. Conexão	97
2.1.1. Probabilidade, Processo e Problemas.	97
2.1.2. Híbridos, Primitivos e Eliminativos	104
2.2. Direção	109
2.3. Seleção	113
Referências	117

(IV) Consciência	127
Robert Van Gulick	

1. História da Questão	128
2. Conceitos da Consciência	131
2.1. A Consciência da Criatura	131
2.2. A Consciência de Estados	133
2.3. Consciência como uma Entidade	135
3. Problema da Consciência	137
4. A Questão Descritiva: <i>Quais</i> são as Características da Consciência?	138
4.1. Dados de Primeira e de Terceira Pessoa	138
4.2. O Caráter Qualitativo	139
4.3. A Estrutura Fenomenal	141
4.4. Subjetividade	141
4.5. Organização Autoperspectivista	142
4.6. Unidade	143
4.7. Intencionalidade e Transparência	144
4.8. Fluxo Dinâmico	146
5. A Questão Explicativa: <i>Como</i> a Consciência Pode Existir?	147
5.1. A Diversidade de Projetos Explicativos	147
5.2. A Lacuna Explicativa	149
5.3. Explicação Reducionista e não-Reducionista	151
5.4. Expectativas de Sucesso Explicativo	153
6. A Questão Funcional: <i>Por Que</i> a Consciência Existe?	154
6.1. Estatuto Causal da Consciência	154

6.2. Controle Flexível	156
6.3. Coordenação Social	157
6.4. Representação Integrada	157
6.5. Acesso Informacional	159
6.6. Livre-Arbitrio	159
6.7. Motivação Intrínseca	160
6.8. Papéis Constitutivos e Contingentes	161
7. Teorias da Consciência	161
8. Teorias Metafísicas da Consciência	162
8.1. Teorias Dualistas	162
8.2. Teorias Ficalistas	164
9. Teorias Específicas da Consciência	168
9.1. Teorias de Ordem Superior	168
9.2. Teorias Representacionalistas	171
9.3. Teorias Cognitivas	173
9.4. Teorias Neurais	175
9.5. Teorias Quânticas	176
9.6. Teorias Não-Físicas	178
10. Conclusão	179
Referências	179
Sobre tradutores e revisores	194

SOBRE A SÉRIE INVESTIGAÇÃO FILOSÓFICA

A *Série Investigação Filosófica* é uma série de livros de traduções de verbetes da Enciclopédia de Filosofia da Stanford (*Stanford Encyclopedia of Philosophy*) e de outras plataformas reconhecidas, que intenciona servir tanto como material didático para os professores das diferentes sub-áreas e níveis da Filosofia quanto como material de estudo para a pesquisa e para concursos da área. Nós, professores, sabemos o quão difícil é encontrar bom material em português para indicarmos. E há certa deficiência na graduação brasileira de filosofia, principalmente em localizações menos favorecidas, com relação ao conhecimento de outras línguas, como o inglês e o francês. Tentamos, então, suprir essa deficiência, ao introduzirmos essas traduções ao público de língua portuguesa, sem nenhuma finalidade comercial e meramente pela glória da filosofia.

Essas traduções foram todas realizadas por filósofos ou por estudantes de filosofia supervisionados e revisadas por especialistas na área. Todas as traduções de verbetes da Stanford foram autorizadas pelo querido Prof. Dr. Edward Zalta, editor da Enciclopédia de Filosofia da Stanford; por isso o agradecemos imensamente. Sua disposição para ajudar brinda os países de língua portuguesa com um material filosófico de excelência, que será para sempre disponibilizado gratuitamente no site da Editora da Universidade Federal de Pelotas (Uditora UFPel) e do NEPFIL/UFPel, dado o nosso maior princípio se fundar na ideia de conhecimento livre e a nossa maior intenção ser o desenvolvimento da filosofia em língua portuguesa e do seu ensino. Aproveitamos o ensejo para agradecer também ao editor da Editora UFPel, na figura do Prof. Dr. Juliano do Carmo, que apoiou nosso projeto desde o início. Agradecemos também a todos os organizadores, tradutores e revisores, que participam de nosso projeto. Sem sua dedicação voluntária, nosso trabalho não teria sido possível. Esperamos, com o início desta coleção, abrir as portas para o crescimento desse projeto de tradução e trabalharmos em conjunto pelo crescimento da filosofia em português.

Prof. Dr. Juliano Santos do Carmo
Prof. Dr. Rodrigo Reis Lastra Cid
Editores da Série Investigação Filosófica

Introdução

Um dos grandes desafios da era da informação consiste em filtrar informações claras, rigorosas e atualizadas sobre tópicos importantes. O mesmo vale para a filosofia. Como encontrar conteúdo filosófico confiável em meio a milhares de artigos publicados diariamente na internet? Para ir ainda mais longe, como encontrar uma introdução a algum tópico com uma lista de referências bibliográficas atualizadas e que seja organizada por um especialista da área? Já que você começou a ler este livro, é provável que tenha ouvido falar em algum momento da Enciclopédia Stanford de Filosofia, disponível gratuitamente na internet desde 1995. A página da Stanford faz exatamente isso. Ela cobre inúmeros verbetes sobre quase tudo em filosofia, oferecendo informação confiável sobre temas como feminismo, filosofia na América Latina, zumbis, metafísica e milhares de outros tópicos. A sugestão típica aos profissionais da filosofia que têm interesse em começar a estudar algum tópico, mas não sabem sequer por onde começar, é procurar pelo verbete relevante da Stanford.

Infelizmente, o grande empecilho da página para o nosso contexto é que todos os verbetes estão disponíveis apenas em inglês, algo que cria um obstáculo a inúmeros estudantes e interessados que não dominam a língua. Neste sentido, acreditamos que esta publicação preencha um hiato importante na filosofia brasileira. Embora a filosofia no Brasil tenha se profissionalizado e expandido ao longo dos anos, há ainda lacunas em diversas áreas, e os verbetes foram aqui selecionados cuidadosamente no intuito de preencher ao menos algumas dessas lacunas. Este livro apresenta quatro verbetes cruciais para a filosofia, aqui divididos em quatro capítulos: (1) condições necessárias e suficientes, (2) referência (3) causalção e (4) consciência.

O primeiro capítulo discute a teoria padrão de condições necessárias e suficientes, mencionando a sua importância para a filosofia em geral (seção 1), incluindo relações com outros capítulos deste livro (por exemplo, os verbetes de referência e causalção). A importância para a filosofia geral se dá principalmente pela relação entre a investigação sobre a natureza das coisas e a obtenção de uma definição explícita dessa natureza. Assim, saber o que é uma definição e

quais os seus problemas possíveis é relevante para o trabalho do profissional da filosofia. Em relação à causação, vale lembrar a influente teoria de J. L. Mackie. De acordo com esta teoria, procura-se definir a noção de uma coisa causando outra por meio da condição INUS (partes insuficientes, mas necessárias, de uma condição que é ela própria não-necessária, mas suficiente para seus efeitos).

Já o tópico da referência (capítulo 2) é igualmente relevante. Houve um tempo em que a filosofia analítica se preocupava majoritariamente com questões sobre a linguagem, principalmente sobre tópicos da referência e do significado. Embora isso tenha de fato mudado (principalmente em virtude da virada metafísica das últimas décadas), é inegável que muitos debates contemporâneos repousam em avanços sobre questões envolvendo referência: Se de fato referimos, qual então o mecanismo da referência? Termos referenciais referem em virtude de quê? Qual o significado de termos referenciais? Há um único mecanismo de referência comum a todos os termos referenciais? O verbete investiga essas questões ao abordar termos referenciais particulares como nomes próprios, descrições definidas, indexicais e demonstrativos. A seção final ainda discute o problema de saber se há de fato a relação de referência.

É claro que, como metafísicos, não poderíamos deixar de incluir um verbete exclusivamente de metafísica: causação. Como teorizar sem causação? É praticamente inconcebível pensar em diversas questões sem o conceito de causa. Pense, por exemplo, nas noções de conhecimento, responsabilidade, livre-arbítrio, percepção e, por incrível que pareça, referência. (A teoria causal da referência, apresentada na seção 2.2 do segundo capítulo, naturalmente invoca o conceito de causação). O verbete de causação é amplo o suficiente para apresentar um bom panorama das discussões contemporâneas sobre o tema – escrito por Jonathan Schaffer, um dos principais especialistas em metafísica contemporânea e defensor da teoria de causação como contraste (ao contrário da teoria padrão em que há apenas dois relata, a teoria de Schaffer estipula quatro relata causais).

Finalmente, o capítulo sobre consciência aborda a questão central sobre a metafísica da mente, centralizando a discussão na natureza da consciência e como ela se relaciona com outros aspectos (não-conscientes) da realidade.

Editamos este volume pela importância dos verbetes para a caixa de ferramentas das filósofas e dos filósofos. Além disso, esses verbetes são cruciais para áreas mais específicas da filosofia, como a metafísica do livre-arbítrio e das leis da natureza. E a relação entre esses verbetes é cristalina. Em relação ao livre-arbítrio, a intuição tradicional é a de que as noções de livre-arbítrio e consciência estão profundamente relacionadas; tipicamente se pensa que a consciência é uma condição necessária para o livre-arbítrio. O mesmo vale para causação. O debate recente em torno do livre-arbítrio envolvendo os cenários de Frankfurt são estruturalmente análogos aos casos de preempção causal, discutidos na seção 2 do capítulo sobre metafísica da causação. Além disso, a teoria contrafactual da causação lá discutida é amiúde utilizada em debates sobre a incompatibilidade entre o livre-arbítrio e o determinismo. O capítulo sobre causação está intimamente relacionado com o debate sobre a metafísica das leis da natureza. Vale a pena notar que essas discussões serão abordadas em edições futuras.

Esperamos que essas traduções possam ser úteis aos estudantes, professores e pesquisadores da área da filosofia em geral.

Prof. Dr. Rodrigo Reis Lastra Cid

Prof. Dr. Pedro Merluzzi

Organizadores do Volume “Textos Seleccionados em Metafísica, Mente e Linguagem”

(I) Condições Necessárias e Suficientes*

Autor: Andrew Brennan

Tradução: Luiz Helvécio Marques Segundo

Revisão: Sérgio R. N. Miranda & Pedro Merluzzi

Uma ferramenta conveniente na busca de definições precisas é a especificação de condições necessárias e/ou suficientes para a aplicação de um termo, do uso de um conceito, ou da ocorrência de algum fenômeno ou evento. Por exemplo, sem água e oxigênio não haveria vida humana; por isso, essas coisas são condições necessárias para a existência dos seres humanos. Os *cockney*, de acordo com a definição tradicional, são todos e apenas aqueles nascidos em locais onde se pode ouvir os sinos da Igreja de St. Mary-le-Bow. Assim, nascer numa área específica é uma condição simultaneamente necessária e suficiente para ser um *cockney*.

Assim como outros conceitos fundamentais, os conceitos de condições necessárias e suficientes não podem ser facilmente especificados em outros termos. Este artigo mostra o quão esquivada é a busca por uma definição dos

* Tradução do verbete "Necessary and Sufficient Conditions" de Andrew Brennan publicado pela *Stanford Encyclopedia of Philosophy* (Edição do Inverno de 2011), edição de Edward N. Zalta, URL = <https://plato.stanford.edu/archives/win2011/entries/necessary-sufficient/>. Publicado com a autorização do editor Prof. Dr. Edward Zalta. A tradução deste verbete foi originalmente publicada em *Investigação Filosófica*: vol. 3, n. 2, artigo digital 1, 2012. Disponível em: <https://periodicos.unifap.br/index.php/investigacaofilosofica/article/view/4866/2192>.

The following is the translation of the entry on Necessary and Sufficient Conditions by Andrew Brennan, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/spr2012/entries/necessary-sufficient/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/necessary-sufficient/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and publish this entry.

termos “necessário” e “suficiente”, indicando a existência de uma ambiguidade sistemática nos conceitos de condições necessárias e suficientes. Também mostra a conexão entre os enigmas sobre essa questão e as questões problemáticas que giram em torno da palavra “se” e o seu uso em frases condicionais.

1. Filosofia e Condições

Uma ambição da filosofia do século vinte foi analisar e refinar as definições dos termos significativos – e os conceitos expressos por eles – na esperança de lançar luz sobre problemas intrigantes como os da verdade, moralidade, conhecimento e existência, que estão fora do alcance de uma solução científica. Central a esse objetivo estava a especificação, pelo menos em parte, de condições a serem satisfeitas para a aplicação correta dos termos, ou sob as quais se poderia dizer que estão realmente presentes em certos fenômenos. Ainda hoje, a única contribuição da filosofia aos estudos interdisciplinares da consciência, da evolução da inteligência, do significado de altruísmo, da natureza da obrigação moral, do escopo da justiça, do conceito de dor, da teoria da percepção e assim por diante, repousa em sua capacidade de fornecer altos graus de rigor e exatidão conceitual aos argumentos nessas áreas.

Se a memória é uma capacidade de localizar as nossas experiências e testemunhos passados, então uma condição necessária para que Penélope se lembre de que deu uma palestra é que ela tenha ocorrido no passado. Ao contrário, Penélope se lembrar da palestra é suficiente para inferir que foi dada no passado. Numa conhecida tentativa de usar a terminologia das condições necessárias e suficientes para definir o que uma coisa tem de ser para ser causa de outra, J. L. Mackie propôs que as causas são no mínimo condições INNS, isto é, “partes Insuficientes embora Necessárias de uma condição que é em si Não-necessária embora Suficiente” para os seus efeitos (Mackie, 1965). O que, então, é uma condição necessária (ou suficiente)? Este artigo mostra que uma precisão completa nessa questão é difícil. Embora a noção de condição suficiente possa ser usada para definir o que é uma condição necessária (e vice-versa), não há maneira clara de oferecer uma abordagem precisa e compreensiva do significado do termo “condição necessária (ou suficiente)”. As advertências de Wittgenstein contra a teorização e a super-generalização prematuras, e o seu *insight* de que muitos termos cotidianos apanham famílias, deveriam despertar cautela na espera de uma especificação completa e não ambígua do que constitui uma condição necessária, ou uma condição suficiente.

2. A Teoria Padrão: Funções de Verdade e Reciprocidade

A porta da frente está fechada. A fim de abri-la (de uma maneira normal, não violenta) e entrar em casa, tenho primeiro de usar a minha chave. Uma condição necessária para se abrir a porta sem violência é, pois, usar a chave. Portanto, parece ser verdade que:

i. Se eu abri a porta, usei a chave.

Podemos usar a interpretação verofuncional do “se” a fim de propor que a consequente de qualquer condicional (em (i), a consequente é “Usei a chave”) especifique uma condição necessária para a verdade da antecedente (em (i), “Abri a porta”)? Muitos textos de lógica e pensamento crítico usam tal abordagem, e por conveniência podemos chamá-la de “a teoria padrão” (veja Blumberg, 1976, pp. 133-4; Hintikka e Backman, 1991, p. 328 para exemplos dessa abordagem).

A teoria padrão faz uso do fato de que na lógica clássica, a função de verdade “ $p \supset q$ ” (“Se p , q ”) é falsa somente quando p é verdadeira e q é falsa. A relação entre “ p ” e “ q ”, nesse caso, é geralmente designada como *implicação material*. Nessa abordagem de “se p , q ”, se a condicional “ $p \supset q$ ” é verdadeira, e p é o caso, então q também é o caso; do mesmo modo, se q não for verdadeira, então p também não pode ser verdadeira (se a condicional inteira for verdadeira). Desse modo, a teoria padrão sustenta que quando a condicional “ $p \supset q$ ” é verdadeira, a verdade da consequente, “ q ”, é necessária para a verdade da antecedente, “ p ”, e a verdade da antecedente é, por sua vez, suficiente para a verdade da consequente. Essa relação entre condições necessárias e suficientes satisfaz a equivalência formal entre a fórmula condicional e a sua contrapositiva (“ $\neg q \supset \neg p$ ” é a *contrapositiva* de “ $p \supset q$ ”). Passando da conversa da verdade das afirmações para falar de estados de coisas, podemos igualmente e de modo correto dizer, de acordo com a teoria padrão, que usar a chave foi uma condição necessária para abrir a porta.

Dada à teoria padrão, as condições necessárias e suficientes são relações conversas, havendo um tipo de espelhamento ou reciprocidade entre as duas: B ser uma condição necessária de A é equivalente a A ser uma condição suficiente de B (e vice-versa). Assim, parece que qualquer frase condicional verofuncional expressa uma condição tanto suficiente quanto necessária. Suponha que se Nellie é um elefante, então tem uma tromba. Ser um elefante é uma condição suficiente para ela ter uma tromba; ter uma tromba é por sua vez

uma condição necessária para Nellie ser um elefante. De fato, a afirmação sobre a condição necessária é simplesmente outro modo de colocar a afirmação sobre a condição suficiente, assim como a contrapositiva de uma fórmula é logicamente equivalente à fórmula original.

É também possível usar “somente se” para identificar uma condição necessária: podemos dizer que Jonas foi engolido por uma baleia somente se foi engolido por um mamífero, pois se uma criatura não é um mamífero, não é uma baleia. A teoria padrão geralmente sustenta que “Se p , q ” e “ p somente se q ” são modos equivalentes de expressar o “ $p \supset q$ ” verofuncional. Equivalente a (i) acima, de acordo com essa abordagem, é a frase “Abri a porta somente se usei a chave” – uma maneira perfeitamente natural de indicar que o uso da chave era necessário para abrir a porta.

A abordagem das condições necessárias e suficientes recém delineada é particularmente conveniente para lidar com condições lógicas. Por exemplo, da verdade de uma conjunção, pode-se inferir que cada uma de suas componentes é verdadeira (se “ p e q ” é verdadeira, então “ p ” é verdadeira e “ q ” é verdadeira). Suponha agora que é verdade que está chovendo e fazendo sol ao mesmo tempo. Essa é uma condição suficiente para “está chovendo” ser verdadeira. Que esteja chovendo é – ao contrário – uma condição necessária para ser verdade que está chovendo e fazendo sol ao mesmo tempo. Uma abordagem similar parece funcionar para contextos conceituais e definicionais. Assim, se o conceito de memória for analisado como o conceito de uma faculdade de localizar eventos ocorridos no passado, o fato de que um evento esteja agora no passado é uma condição necessária para que eu me lembre dele no presente. Se a água é quimicamente definida como um líquido constituído principalmente de H_2O , então se um copo contém água, ele contém principalmente H_2O . O copo conter H_2O é uma condição necessária para que contenha água.

A despeito de seu apelo inicial, objeções à teoria padrão têm sido feita por teóricos de várias perspectivas. De maneira sumária, as objeções repousam na ideia de que o “se” em português nem sempre expressa um tipo uniforme de condição. Se diferentes tipos de condições são expressas pela palavra “se”, argumentam os objetores, então seria sensato descobri-los antes de se começar a tentar formalizar e sistematizar os conceitos de *necessário* e *suficiente*. Ao tentar mostrar que há uma ambiguidade contaminando as frases do português introduzidas por “se”, os críticos têm se focado em duas doutrinas que consideram erradas: em primeiro lugar, que há uma reciprocidade entre condições necessárias e suficientes, e, em segundo, que “se p , q ” e “ p somente

se p ” são modos equivalentes de dizer a mesma coisa.

3. Problemas para a Teoria Padrão

Dadas duas frases verdadeiras quaisquer A e B , a condicional “Se A , então B ” é verdadeira. Por exemplo, dada a verdade de que o sol é feito de gás e também de que os elefantes têm quatro patas, então a condicional verofuncional “Se os elefantes têm quatro patas, então o sol é feito de gás” é também verdadeira. No entanto, a natureza gasosa do sol normalmente não seria considerada como uma condição conceitualmente, ou mesmo contingentemente, necessária para a quadrupedidade dos elefantes. Na verdade, de acordo com a teoria padrão, qualquer verdade será uma condição necessária para a verdade de toda e qualquer afirmação, e qualquer falsidade será uma condição suficiente para a verdade de qualquer afirmação que queiramos considerar.

Esses resultados estranhos não surgiriam nas lógicas não-clássicas em que se exige que as premissas sejam relevantes para as conclusões delas tiradas, e que as antecedentes das condicionais verdadeiras sejam do mesmo modo relevantes para as consequentes. Mas mesmo nas versões das lógicas relevantes que evitam alguns desses resultados estranhos, é difícil evitar todos os chamados “paradoxos da implicação”. Por exemplo, uma contradição (uma asserção da forma “ p e não- p ”) será uma condição suficiente para a verdade de qualquer afirmação a menos que a semântica para a lógica em questão permita a inclusão de mundos inconsistentes (para mais detalhes, veja o verbete lógica: relevância, e para uma explicação da relevância em termos da ideia de “restrição de significado” veja o capítulo 1 de Brady, 2006).

Essas estranhezas poderiam ser rejeitadas como meras anomalias não fosse pelo fato de os autores terem identificado aparentemente diversos outros problemas associados com as ideias de reciprocidade e equivalência mencionadas no fim da seção anterior. De acordo com a teoria padrão, há um tipo de reciprocidade entre as condições necessárias e suficientes, e as frases “se p , q ” podem ser sempre parafraseadas por frases “ p somente se q ”. No entanto, como os autores da linguística observaram, nenhuma dessas afirmações satisfaz ou a compreensão mais natural das condições necessárias (e suficientes), ou do comportamento do “se” (e do “somente se”) no português. Considere, por exemplo, o seguinte caso (tirado de McCawley, 1993, p. 317):

ii. Se você encostar em mim, gritarei.

No caso da porta, enquanto que usar a chave era necessário para abri-

la, nenhuma afirmação paralela parece funcionar para (ii): na leitura natural dessa afirmação, o meu grito não é necessário para que você encoste em mim. McCawley afirma que a oração subordinada introduzida por “se” no português padrão fornece a condição – seja epistêmica, temporal ou causal – para a verdade da oração principal. A interpretação natural de (ii) é que o meu grito depende de você encostar em mim. Considerar o meu grito como uma condição necessária para que você me toque parece inverter as dependências. Uma questão similar surge quando se sustenta que (ii) acarreta que você me tocará só se eu gritar.

Uma falha similar de reciprocidade ou espelhamento surge no caso do exemplo da porta ((i) acima). Enquanto que abrir a portar dependeu, temporalmente ou causalmente, de primeiro se usar a chave, seria errado pensar que usar a chave dependeu, ou temporalmente ou causalmente, de abrir a porta. Assim, que tipo de condição a antecedente expressa? Para deixar isso claro, podemos considerar um par desconcertante de frases condicionais (uma modificação de Sanford, 1989, pp. 175-6):

iii. Se Lambert aprender a tocar, comprarei para ele um violoncelo.

iv. Lambert aprende a tocar somente se eu comprar para ele um violoncelo.

Note que essas duas afirmações não são equivalentes em significado, muito embora os manuais tratem comumente “se p , q ” apenas como outro modo de dizer “ p somente se q ”. Ao passo que (iii) expressa uma condição na qual compro um violoncelo para Lambert (presumivelmente ele aprende primeiro usando um violoncelo emprestado, ou talvez alugado), (iv) expressa uma condição necessária de Lambert aprendendo a tocar o instrumento primeiro (pode haver outras também). Na verdade, se tomarmos as duas conjuntamente, as afirmações deixam o pobre Lambert sem expectativas de ganhar o violoncelo de mim. Se (iv) fosse equivalente a (iii), combinar as duas afirmações não conduziria a tal *impasse*.

Mas como podemos formular (iii) em termos de “somente se”? Um equivalente natural português é surpreendentemente difícil de formular. Talvez fosse algo assim:

v. Lambert aprendeu a tocar violoncelo somente se comprei para ele um violoncelo.

Onde o passado (“aprendeu”/ “comprei”) foi usado para tentar manter as dependência em ordem. Contudo, (v) não está completamente correta, pois pode

ser lida como implicando que o sucesso de Lambert é dependente de eu ter primeiro comprado o violoncelo para ele – algo que certamente não foi implicado por (iii). Uma versão ainda melhor (embora não completamente satisfatória) requer ajuste adicional, digamos:

vi. Lambert terá aprendido a tocar violoncelo somente se comprei para ele um violoncelo.

Dessa vez, não é tão fácil interpretar (vi) como implicando que comprei um violoncelo para Lambert antes dele ter aprendido a tocar. Essas mudanças no tempo verbal conduziram alguns autores a argumentar que as condicionais em português envolvem quantificação implícita no tempo (veja, por exemplo, von Fintel, 1988). A avaliação dessa tese vai além do escopo do presente artigo (veja a entrada sobre condicionais e a discussão detalhada em Bennett, 2003).

O que o argumento parece sugerir é que diferentes tipos de dependência são expressas pelo uso da construção condicional: (iv) não é equivalente a (iii) porque a condição necessária em (iii) (que compro um violoncelo para Lambert) fornece aquilo que poderia ser chamado de uma *razão para pensar* que Lambert aprendeu a tocar violoncelo. Em contraste, a mesma condição – que compro um violoncelo para Lambert – parece cumprir uma função diferente de acordo com (iv) (designadamente, que primeiro tenho de comprar um violoncelo para ele antes dele aprender a tocar). Na seção seguinte, a possibilidade de distinguir entre tipos diferentes de condições é discutida. Se conseguirmos fazer tais distinções, então isso pode sugerir que há uma ambiguidade sistemática nos conceitos de *condição necessária* (e *suficiente*).

A possibilidade da ambiguidade nesses conceitos levanta outro problema para a teoria padrão. De acordo com ela – como von Wright apontou (von Wright, 1974, p. 7) – as noções de condição necessária e condição suficiente são interdefiníveis:

A é uma condição suficiente de $B =_{df}$ a falta de A é uma condição necessária para a falta de B .

B é uma condição necessária de $A =_{df}$ a falta de B é uma condição suficiente da falta de A .

A ambiguidade ameaçaria essa nítida interdefinibilidade. Na seção seguinte, exploraremos se há algo com que se preocupar aqui. A possibilidade de tal ambiguidade foi explorada no trabalho de Downing (1959; 1975) e Wilson (1979), e tem sido levantada mais recentemente em Goldstein *et al.* (2005), cap. 6. Esses autores têm argumentado que em alguns contextos há uma falta de

reciprocidade entre as condições necessárias e suficientes interpretadas de certo modo. Em outras situações, as condições se relacionam reciprocamente umas às outras do modo exigido pela teoria padrão. Se essas críticas estiverem corretas, e a ambiguidade estiver presente, então não há uma conclusão geral que possa ser estabelecida sobre a reciprocidade, ou a falta dela, entre condições necessárias e suficientes, mas ao invés, haverá uma necessidade de distinguir o sentido da condição que está sendo invocada num contexto particular. Sem a especificação do significado e do contexto, seria também errado fazer a afirmação geral de que frases como “se p , q ” são geralmente parafraseáveis como “ p somente se q ”.

4. Inferências, Razões Para Pensar, e Razões Pelas Quais

As duas afirmações seguintes são equivalentes? (veja Wertheimer, 1968, pp. 3634):

vii. A ocorrência de uma batalha naval amanhã é uma condição necessária e suficiente para a verdade, hoje, de “Haverá uma batalha naval amanhã”.

viii. A verdade, hoje, de “Haverá uma batalha naval amanhã” é uma condição necessária para a ocorrência de uma batalha naval amanhã.

Sanford argumenta que ao passo que (vii) é sensata, (viii) “inverte as coisas” (Sanford, 1989, pp. 176-7). Ele escreve: “a asserção sobre a batalha, se verdadeira, é verdadeira por causa da ocorrência da batalha. A batalha não ocorre por causa da verdade da afirmação” (*ibid.*). O que ele provavelmente quer dizer é que a ocorrência da batalha explica a verdade da afirmação, e não o contrário. Com certeza, as pessoas às vezes comprometem-se apenas para assegurar que aquilo que disseram torne-se verdadeiro; assim, há casos em que a verdade da afirmação explica a ocorrência de um evento. Mas isso parece uma interpretação pouco provável do caso da batalha naval.

Seja S a frase “Haverá uma batalha naval amanhã”. Se S é verdadeira hoje, é correto inferir que uma batalha naval ocorrerá amanhã. Isto é, muito embora a verdade da frase não explique a ocorrência da batalha, o fato dela ser verdadeira permite a inferência da ocorrência do evento. Ascendendo ao modo puramente formal (no sentido de Carnap), podemos resolver o problema limitando explicitamente as relações das inferências àquelas que ocorrem entre as frases ou outros itens que podem ter valores de verdade. É perfeitamente adequado inferir da verdade de S hoje que alguma outra frase é verdadeira

amanhã, como, por exemplo, “há uma batalha naval hoje”. Uma vez que “há uma batalha naval hoje” é verdadeira amanhã se e somente se houver uma batalha naval amanhã, então podemos inferir que do fato de que S é verdadeira hoje, que uma batalha naval ocorrerá amanhã.

Dessa observação, pareceria que há um hiato entre o que é verdadeiro em relação às inferências, e o que é verdadeiro em relação às explicações. Há um sentido (inferencial) em que a verdade de S é tanto uma condição necessária quanto suficiente para a ocorrência da batalha naval. No entanto, há um sentido (explicativo) em que a ocorrência da batalha naval é necessária e suficiente para a verdade de S, mas não vice-versa. Pareceria então que nos casos como (vii) e (viii) haveria inferências que correriam em ambas as direções, enquanto as explicações correriam apenas para um lado. Interpretar (vii) como equivalente a (viii) dependerá do sentido em que as noções de condições necessárias e suficientes estão sendo empregadas.

É possível generalizar essa descoberta? Retornando ao próprio exemplo, o fato de que usei a chave explica por que consegui abrir a porta. Que abra a porta sem usar a força fornece base para se inferir que usei a chave. Eis outro exemplo de McCawley:

ix. Se John vencer a corrida, comemoraremos.

A vitória de John é uma condição suficiente para que tenhamos uma comemoração, e a sua vitória é a razão pela qual estaremos a comemorar. A nossa comemoração, no entanto, certamente não é a razão pela qual ele vence a corrida. Em que sentido, pois, a comemoração é uma condição necessária da vitória de John? Novamente, há uma base para a inferência: nós não comemorarmos é uma base para inferir que John não venceu a corrida. O uso do tempo verbal é sensível à assimetria exposta na seção anterior. O modo natural de se escrever a contrapositiva de (ix) não é “Se não comemoramos, então John não vence a corrida”, mas antes, algo como:

x. Se não comemoramos, John não venceu a corrida.

Ou

xi. Se não estamos comemorando, John não venceu a corrida.

Ou ainda

xii. Se não comemoramos, John não pode ter vencido a corrida.

A reciprocidade inferencial e não-reciprocidade explicativa, no caso das condicionais, parece não ser diferente dos casos das equações lógicas e

matemáticas em geral. Por exemplo, a identidade clássica de Newton, $f=ma$, pode ser reescrita nas formas equivalentes como $m=f/a$ ou $f/a=m$. Do ponto de vista algébrico, todas essas formas expressam a mesma coisa. Ora, suponhas que força é a medida daquilo que leva uma partícula a um certo estado. Assim, diríamos que ao passo que a força causa a aceleração, a razão f/a não causa ou explica a massa, muito embora a determine (veja o *Epílogo* de Pearl, 2000 para uma tentativa não-técnica de tratar a representação da intervenção causal pelas notações algébricas).

Parece haver pelo menos três relações diferentes que podemos distinguir em conexão com as afirmações condicionais, cada uma delas ligadas às questões de necessidade e suficiência. A primeira é a relação de implicação simbolizada pelo operador gancho, “ \supset ”, ou talvez por algum operador relevante de implicação. Como já notado, esse operador capta algumas relações ambigüidade. Por exemplo, vimos que da verdade de uma conjunção pode-se inferir que cada uma das suas componentes é verdadeira (de “ p e q ” podemos inferir que “ p ” é verdadeira e que “ q ” é verdadeira). O gancho, ou o operador relevante de implicação, parece captar uma das relações encontradas no caso da batalha naval, que se pode pensar como ocorrendo paradigmaticamente entre portadores de verdade, embora se possa pensar, de modo mais solto, em termos de estados de coisas. Por causa dessa relação conseguimos manter a tese da reciprocidade da teoria padrão.

No entanto, duas relações extras estão geralmente contidas nas reflexões sobre condições necessárias e suficientes. Para identificá-las, considere as coisas diferentes que se pode querer dizer como:

xiii. Se Lambert esteve presente, foi um bom curso.

Um cenário em termos do qual (xiii) pode ser entendida é aquele no qual Lambert é invariavelmente um colaborador ativo em qualquer curso que participa. Além do mais, as suas contribuições são sempre boas, garantindo, assim, satisfação para todos os que participaram do curso. Nesse caso, a presença de Lambert foi a *razão pela qual* o curso o seminário foi bom. Um cenário diferente descreve Lambert como alguém sendo quase que infalível para encontrar cursos que vão ser bons, muito embora ele nem sempre seja ativo na discussão. A presença de Lambert num curso, de acordo com esse relato, fornece uma *razão para pensar* que o curso vai ser bom. Poderíamos dizer que de acordo com o primeiro relato, o seminário é bom porque Lambert está nele. No segundo caso, Lambert está nele porque o seminário é bom. Exemplos desse

tipo foram primeiro introduzidos em Wilson (1979) inspirados pelo trabalho de Peter Downing (1959; 1975). Note que o gancho (como entendido na lógica clássica) não capta a *razão para pensar* na relação, pois permite que qualquer verdade seja inferida de qualquer outra afirmação.

As condições que expressam uma *razão pela qual* e as que expressam uma *razão para se pensar que* podem lançar luz sobre algumas peculiaridades encontradas anteriormente. Eu abrir a porta é uma razão para se pensar que usei a chave, e não uma razão pela qual a usei. Ele aprender a tocar o instrumento é a razão pela qual (no caso (iii)) comprei um violoncelo, e eu comprar um violoncelo para ele é (no mesmo caso) uma razão para se pensar que – mas não uma razão pela qual – ele aprendeu a tocar o instrumento. A nossa comemoração é uma razão para se pensar que John venceu a corrida no caso (ix), mas não uma razão pela qual ele venceu.

Embora haja às vezes uma correlação entre as razões pelas quais, por um lado, e as razões para se pensar, por outro, nenhuma generalização pode ser feita seguramente (embora Wilson (1979) ofereça diversas sugestões sobre as conexões entre essas noções). Se *A* é uma razão pela qual *B* ocorreu (e assim talvez também uma razão para se pensar que *B* ocorreu), então a ocorrência de *B* será às vezes uma razão para se pensar que *A* ocorreu – embora não uma garantia. Se *A* não é nada além do que uma razão para se pensar que *B* ocorreu, então *B* será às vezes uma razão para se pensar por que *A* ocorreu – embora não uma garantia. Voltando ao nosso exemplo inicial, eu ter aberto a porta sem violência foi uma razão para se pensar que usei a chave. Eu ter usado a chave, no entanto, não foi apenas uma razão para se pensar que abri a porta, mas uma das razões pelas quais consegui abrir a porta. O que é importante é que a oração subordinada de uma condicional pode fazer quaisquer das três coisas descritas na presente seção: (i) introduzir uma premissa cuja consequente se segue do modo determinado por um operador como o gancho; (ii) expressar uma razão pela qual aquilo que é expresso na consequente seja o caso; (iii) expressar uma razão para se pensar que aquilo que é expresso pela consequente é o caso (embora não expresse uma razão pela qual ela seja o caso).

Em geral, se a explicação é direcional, pode não parecer surpreendente que quando *A* explica *B*, não seja geralmente o caso que *B*, ou a sua negação, seja por sua vez uma explicação de *A* (ou a sua negação). A vitória de John explica a nossa comemoração, mas o fato de não comemorarmos não é (normalmente) uma explicação plausível de ele não ter vencido. A presença de

Lambert pode explicar por que o curso teve grande sucesso, mas um curso entediante não é – em qualquer conjunto normal de circunstâncias – uma razão pela qual Lambert não está lá. Esse resultado coloca em causa o entendimento comum de que se *A* é uma condição suficiente de *B*, tipicamente será o caso que *B* é uma condição necessária para *A*, e a falsidade de *B* uma condição suficiente para a falsidade de *A*.

A fim de reduzir a ameaça imposta pelo hiato entre as condições inferenciais e explicativas, podem algumas condicionais serem entendidas em termos de ambos os papéis? Por exemplo, não pode ser o caso que a presença de Lambert no seminário seja tanto uma razão para se pensar que o seminário será bom, e – como em nosso primeiro relato – também uma razão pela qual o seminário será bom? É como se houvesse uma forma fraca de contraposição que valesse nesse caso (essa sugestão é feita em Wilson (1979)): o curso não ser bom fornece pelo menos uma razão para se pensar que Lambert não estava presente. Ora, se essa forma fraca de contraposição se mantém, então as condições necessárias e suficientes podem ser afinal consideradas como conversas. Não é claro que a defesa da contraposição será bem sucedida. Pois não há reciprocidade explicativa no caso dado, apenas uma reciprocidade entre a alegação de que se *A* é uma razão para se pensar que *B*, então a ausência de *B* é uma razão para se pensar que *A* também está ausente. Em outras palavras, a forma supostamente “fraca” da contraposição não é na verdade um caso de contraposição. De fato, no primeiro relato que oferecemos, foram a presença de Lambert, as suas contribuições ativas, o seu forte bom humor, etc., os fatores que explicam o sucesso dos cursos que ele participa. Nesse relato, não há acarretamento da conversa – a ausência de Lambert de qualquer curso particular não explicaria de maneira alguma ou forneceria razões pelas quais o curso não foi bom.

Poder-se-ia tentar argumentar também que, a despeito da direcionalidade das afirmações causais, há pelo menos uma forma fraca de contraposição que vale para algumas condicionais causais. Num artigo recente (Gomes, 2009), Gilberto Gomes propõe que quando se diz que "*A*" é uma condição causalmente suficiente para "*B*", ou "*B*" uma condição causalmente suficiente de "*A*", então alguma forma de reciprocidade entre os dois tipos de condições se sustenta, e, por isso, alguma versão da contraposição será válida. Voltando ao exemplo (ii), suponha que a interpretemos como expressando uma condição causal – que ao encostar em mim você causaria o meu grito. Gomes sugere que "*A*" denota uma causa suficiente de "*B*" desde que (1) "*A*" especifique

a ocorrência de um evento que causaria outro evento "B", e que faça isso (2) estabelecendo uma condição cuja verdade é suficiente para inferir a verdade de "B". Nesse caso, poderíamos sustentar ainda que "B", por sua vez, denota um efeito necessário de "A", querendo dizer então que a verdade de "B" fornece uma condição necessária para a verdade de "A" (Gomes, 2009, pp. 377-9). Se essa proposta se sai melhor do que o apelo à contraposição fraca discutida acima é algo a se pensar.

Mesmo que seja possível distinguir esses diferentes papéis que a oração subordinada pode desempenhar (haverá outros também), não é sempre fácil isolá-los em cada caso. O apelo às "razões pelas quais" e às "razões para pensar" nos permite identificar aquilo que parece ser as ambiguidades no "se" e na terminologia das condições necessárias e suficientes. Infelizmente, o próprio conceito de *explicação* é demasiado vago para ser útil aqui, pois podemos explicar um fenômeno citando uma razão para se pensar que tal fenômeno foi o caso, ou citando uma razão pela qual tal fenômeno é o caso. Outros problemas surgem quando consideramos leis matemáticas, físicas, ou outras leis que podem estar envolvidas no uso de conceitos (um *locus classicus* para essas questões é Sellars (1948)). Por exemplo, a verdade de "esta figura é um polígono" é suficiente para inferir "a soma dos ângulos externos desta figura é 360 graus". De "a soma dos ângulos externos da figura não é 360 graus" podemos inferir "a figura não é um polígono". Essas inferências não são logicamente triviais. Antes, dependem de definições geométricas e princípios matemáticos, e, por isso, esse é um caso de condições matematicamente necessárias e suficientes. Mas, ao dizer isso, abrimos caminho para se reconhecer que os resultados matemáticos também nos dão pelo menos uma razão para pensar o seguinte: porque uma figura é um polígono, os seus ângulos externos somarão 360 graus. Podemos ainda ser capazes de pensar em contextos que o fato de uma figura ser um polígono fornece uma razão pela qual os seus ângulos externos somem 360 graus.

Um aspecto similar acontece com a teoria do conhecimento em que geralmente se assume que se sei que p , então p é verdadeira. A verdade de p é uma condição necessária para saber que p de acordo com tais abordagens. Ao dizer isso, não excluimos as alegações mais fortes do que simplesmente dizer que a verdade de p se segue do fato de que sabemos que p . Uma crença ser verdadeira – por exemplo – pode ser (parte de) uma razão para se pensar que essa crença constitui conhecimento. Outros casos envolvem inferências permitidas pela física, pela biologia, e pelas ciências naturais – inferências que

envolverão condições causais ou nômicas. Novamente é preciso ter cuidado ao determinar se é a relação da *razão pela qual* ou a da *razão para acreditar* que está sendo estabelecida. O aumento da média de energia cinética de suas moléculas não apenas implica que a temperatura de um gás está subindo, mas também fornece uma razão pela qual a temperatura está aumentando. No entanto, se a temperatura é apenas um modo de medir a energia cinética molecular média, então uma mudança na temperatura será uma razão para se pensar que a energia cinética média das moléculas mudou, e não uma razão pela qual mudou.

Como mencionado no início do artigo, a especificação de condições necessárias e suficientes tradicionalmente tem sido parte do trabalho dos filósofos na análise dos termos, conceitos e fenômenos. As investigações filosóficas do conhecimento, da verdade, da causalidade, da consciência, da memória, da justiça, do altruísmo e de uma gama de outras questões não visam estabelecer relações explicativas, mas, antes, identificar e desenvolver relações conceituais (veja Jackson (1998) para uma abordagem detalhada da análise conceitual). As análises filosóficas não oferecem geralmente condições que fornecem *razões pelas quais* ou *razões para acreditar que*. Nem fornecem definições de dicionários. Que Nellie seja um elefante não é uma (ou a) razão pela qual ela é um animal, não mais do que uma figura que é um quadrado é uma razão pela qual tem quatro lados. Antes, os elefantes são um tipo de animal, e quadrados são um tipo de figuras de quatro lados. Especificar as condições necessárias para a verdade da frase “esta figura é um quadrado” é especificar várias condições, incluindo “esta figura tem quatro lados”, “esta figura é plana”, e “esta figura é fechada”. Se quaisquer uma dessas condições for falsa, então a frase “esta figura é um quadrado” é também falsa. Conversamente, a verdade de “esta figura é um quadrado” é uma condição suficiente para a verdade de “esta figura é fechada”. As relações inferenciais nesse caso são determinadas em alguma medida – se bem que inadequadamente, como notado anteriormente – por um operador como o gancho.

Considere agora o nosso exemplo anterior – o da memória. Penélope se lembrar de algo – de acordo com a abordagem padrão da memória – significa (dentre outras coisas) que a coisa lembrada esteve no passado, e que alguns episódios prévios envolvendo Penélope desempenham um papel causal apropriado na sua lembrança atual da coisa em questão. Seria um erro inferir do papel causal de algum episódio passado na lembrança atual de Penélope que a definição de memória envolve condições que sejam explicativas no sentido das

razões pelas quais. Penélope lembrar agora de algum evento não é uma razão pela qual esse evento esteja no passado. Antes, os tratamentos filosóficos da memória procuram por condições que são constituintes *a priori* da verdade de frases como “Penélope se lembra de fazer X”. A exposição de tais condições não explica Penélope estar agora se lembrando de coisas, mas, antes, fornece um *insight* para saber se, e como, “lembrar” tem de ser definido. As condições que expressam uma *razão pela qual* e as que expressam *razão para pensar que* não desempenham um papel nessa parte do trabalho dos filósofos, o que vem a calhar, dado que uma abordagem propriamente formalizada dessas duas relações ainda espera por desenvolvimento.

Finalmente, dever-se-ia notar que nem todas as frases condicionais visam primariamente oferecer condições necessárias e/ou suficientes. Um caso comum envolve aquilo que poderíamos chamar de *condicionais jocosas*. Um amigo de Lis se refere erroneamente à “*Crítica da Razão Pura* de Platão” e Lis observa, “Se Platão escreveu a *Crítica da Razão Pura*, então eu sou Aristóteles”. Ao invés de especificar condições, Lis está empregando uma forma de *redução ao absurdo*. Uma vez que é óbvio que ela não é Aristóteles, a sua piada convida o interlocutor a inferir (por contraposição) que Platão não escreveu a *Crítica da Razão Pura*.

5. Conclusão

Dados os papéis diferentes para o “se” já identificados, dificilmente surpreende que as generalizações sobre as condições necessárias e/ou suficientes são difíceis de formular. Suponha, por exemplo, que alguém tenta expressar uma condição suficiente para um curso ser bom num contexto em que o falante e todos os ouvintes compartilham o ponto de vista de que a presença de Lambert é uma razão pela qual os cursos seriam bons. Nesse caso, poder-se-ia dizer que a presença de Lambert é uma condição suficiente do curso ser bom no sentido de que a sua presença é uma razão pela qual o curso é bom. Ora, há um sentido similar em que a bondade do curso seja uma condição necessária da presença de Lambert? A resposta negativa a essa pergunta já é evidente pela discussão anterior. Se seguimos a proposta de von Wright mencionada anteriormente, temos o seguinte resultado: o curso não ser bom é uma condição suficiente de Lambert não estar presente. Mas isso não pode ser adequadamente interpretado como uma condição suficiente de acordo com qualquer sentido de uma razão pela qual. No máximo, o fato de o curso não ter sido bom pode ser uma razão para se pensar que Lambert não estava lá. Assim,

em geral, como podemos dizer qual tipo de condição está sendo expressa na oração subordinada de uma condicional? Como notado no caso da batalha naval, quando reescrever no modo formal capta o sentido do que esta sendo dito, e quando as formulações “se p , q ” e “ p somente se q ” parecerem idiomáticamente equivalentes, então uma interpretação inferencial será adequada, as equivalências de von Wright se manterão, e a condicional material poderá dar conta de tais casos.

Com já notamos, mesmo o uso inferencial do “se” não é sempre associado primariamente com a tarefa de estabelecer condições necessárias e suficientes. Essa observação, junto com os casos e as distinções introduzidas no presente artigo, mostra a necessidade de se ter cuidado quando passamos das condicionais da linguagem natural à sua análise em termos de condições necessárias e suficientes, e também a necessidade de se ter cuidado ao determinar as últimas condições por meio dos operadores lógicos. Alternativamente, temos de reconhecer que há vários tipos de condicionais, e vários tipos de condições. Assim, embora possamos às vezes usar – e usamos – as afirmações condicionais para expressar condições necessárias e suficientes, e possamos explicar as condições necessárias e suficientes pela análise de alguns dos papéis do “se” nas condicionais da linguagem natural, isso não nos dá tudo o que queremos. Em particular, não parece haver um esquema formal para traduzir as condicionais usadas na linguagem natural em afirmações de um tipo particular qualquer de condição ou vice-versa.

Referências

Blumberg, A. E., 1976. *Logic: A First Course*, New York: Alfred E. Knopf.

Bennett, J., 2003. *A Philosophical Guide to Conditionals*, Oxford: Oxford University Press.

Brady, Ross, 2006. *Universal Logic*, Stanford: CSLI Publications.

Downing, Peter, 1959. “Subjunctive Conditionals, Time Order and Causation”, *Proceedings of the Aristotelian Society*, 59: 126–40.

Downing, Peter, 1975 “Conditionals, Impossibilities and Material Implications”, *Analysis*, 35: 84–91.

Gomes, Gilberto, 2009. “Are Necessary and Sufficient Conditions Converse Relations?”, *Australasian Journal of Philosophy*, 87: 375–87.

Goldstein, L., Brennan, A., Deutsch, M. and Lau, J., 2005. *Logic: Key Concepts in*

- Philosophy, London: Continuum.
- Hintikka, J. and Bachman, J., 1991. *What If ...? Toward Excellence in Reasoning*, London: Mayfield.
- Jackson, F., 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Oxford University Press.
- Mackie, J. L., 1965. "Causes and Conditions", *American Philosophical Quarterly*, 12: 245–65.
- McLaughlin, Brian, 1990. *On the Logic of Ordinary Conditionals*, Buffalo, NY: SUNY Press.
- Sanford, David H., 1989. *If P, then Q: Conditionals and the Foundations of Reasoning*, London: Routledge.
- Pearl, Judea, 2000. *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.
- Von Fintel, Kai, 1997. "Bare Plurals, Bare Conditionals and Only", *Journal of Semantics*, 14: 1–56.
- Sellars, Wilfrid, 1948. "Concepts as involving laws and inconceivable without them", *Philosophy of Science*, 15: 289–315.
- McCawley, James, 1993. *Everything that Linguists have Always Wanted to Know about Logic** (Subtitle: *But Were Ashamed to Ask), Chicago: Chicago University Press.
- Wertheimer, R., 1968. "Conditions", *Journal of Philosophy*, 65: 355–64.
- Von Wright, G. H., 1974. *Causality and Determinism*, New York: Columbia University Press.
- Wilson, Ian R., 1979. "Explanatory and Inferential Conditionals", *Philosophical Studies*, 35: 269–78.
- Woods, M., Wiggins, D. and Edgington D. (eds.), 1997. *Conditionals*, Oxford: Clarendon Press.

(II) Referência*

Autores: Eliot Michaelson & Marga Reimer

Tradução: Lucas Roisenberg Rodrigues

Revisão: Jerzy André Brzozowski & Pedro Merluzzi

A referência é uma relação entre certos tipos de instâncias representacionais e objetos¹. Por exemplo, quando afirmo que “Barack Obama é um democrata”, uso um tipo particular de instância representacional – i.e., o nome “Barack Obama” – o qual se refere a um indivíduo particular – i.e., Barack Obama. Enquanto nomes e outros termos referenciais dificilmente são os únicos

*

Tradução do verbete "Reference" de Eliot Michaelson & Marga Reimer publicado pela *Stanford Encyclopedia of Philosophy* (Edição da Primavera de 2019), edição de Edward N. Zalta, URL = <<https://plato.stanford.edu/archives/spr2019/entries/reference/>>. Publicado com a autorização do editor Prof. Dr. Edward Zalta.

The following is the translation of the entry on Reference by Eliot Michaelson & Marga Reimer, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/spr2019/entries/reference/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/reference/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and publish this entry.

¹ O termo *token* foi aqui traduzido por "instância". De forma similar, traduzimos *representational token* por "instância representacional". Salvo indicação em contrário, "instância" será sempre a tradução do *token*. *Token*, como termo técnico em filosofia e linguística, normalmente significa as instâncias concretas e particulares de um certo tipo. Assim, por exemplo, na palavra "casa" existem quatro letras-instância, isto é, o tipo letra é exemplificado ou instanciado quatro vezes. Todavia, se considerarmos *tipos* de palavras, encontramos a ocorrência de apenas três letras-tipo: C, A e S. Se eu profiro a sentença "a casa é apenas uma casa" temos duas instâncias do substantivo "casa" (aqui considerado como tipo): a primeira instância correspondente ao primeiro proferimento, e a segunda instância correspondente ao segundo proferimento. (Nota do tradutor).

tipos de instâncias representacionais capazes de referir (considere, por exemplo, conceitos, mapas mentais e imagens), instâncias linguísticas como estas há muito tempo estão no centro das investigações filosóficas sobre a natureza da referência. Em conformidade com isto, este verbete se concentrará principalmente na referência linguística².

Assumindo que pelo menos algumas instâncias de expressões linguísticas realmente referem, muitas questões interessantes surgem. Como, por exemplo, a referência linguística se relaciona com o ato de referir – algo que nós, como falantes, fazemos com os termos referenciais? Como exatamente os termos referenciais vêm a referir? Ou seja, em virtude do quê eles referem ao que referem? Existe uma resposta única a esta questão, um único mecanismo de referência, ou diferentes respostas, dependendo do tipo de termo em questão? E qual é exatamente a relação entre a referência e o significado?

As respostas a essas várias questões acabarão por se mostrar intimamente relacionadas; a tarefa deste verbete será traçar alguns dos principais grupos de respostas.

1. Introdução

Utilizamos a linguagem para falar sobre o mundo. Muito do que dizemos parece significativo; uma parte disto, presumivelmente, é até mesmo verdadeira. Por exemplo, eu pareço estar dizendo algo verdadeiro quando, no contexto apropriado, eu assertivamente profiro³:

1-Barack Obama é um democrata.

Como conseguimos realizar tais coisas? Como, por exemplo, eu consigo falar acerca de Barack Obama, e assim dizer coisas verdadeiras e significativas a respeito dele? Em uma palavra: como eu consigo me referir a Barack Obama por meio do nome "Barack Obama"? Metaforicamente, nós parecemos capazes de usar a linguagem para falar sobre mundo porque nossas palavras

² Para mais sobre a referência dos estados mentais, consultar os verbetes sobre representação mental, teorias causais do conteúdo mental, externalismo sobre conteúdo mental, e as teorias teleológicas do conteúdo mental. Para mais sobre a referência de imagens, consultar o verbete sobre a estética de Goodman.

³ No original, *I assertively utter*. O termo *utter* sempre será traduzido por "proferimento". (Nota do tradutor).

são capazes de "engatar" nas coisas do mundo, coisas tais como Barack Obama. Nomes próprios – isto é, expressões como "Barack Obama" e "David Cameron" – são amplamente consideradas como expressões referenciais paradigmáticas. Embora possa parecer implausível a suposição de que todas as palavras referem, que todas as palavras de alguma forma "engatam" em porções da realidade, certas palavras provavelmente são do tipo referencial. Estas incluem: nomes próprios, pronomes, indexicais, demonstrativos, termos para tipos naturais, e muitos outros tipos de termos de propriedade. Descrições definidas são outro, ainda que muito controverso, candidato. Aqui, nos concentraremos em apenas um subconjunto destas – a saber, os frequentemente denominados termos referenciais "singulares". Estes são termos que, supondo que refiram, se referem a objetos e indivíduos, em oposição a grupos ou propriedades. Ainda que os termos referenciais singulares tenham sido submetidos a um intenso escrutínio filosófico nos últimos cem anos, ainda há muito o que discutir.

Este artigo se concentrará em cinco questões centrais relacionadas a respeito da referência: (i) como a relação referencial, que ocorre entre as instâncias de uso⁴ e os objetos, relaciona-se com o ato de referir, algo que os falantes fazem com esses termos? Em outras palavras, os termos se referem ao que eles se referem em virtude de nós os usarmos com tal finalidade, ou nós usamos os termos em virtude de eles já referirem previamente? (ii) Qual é o significado dos termos referenciais? Eles são meramente referenciais, ou fazem algo mais? (iii) Qual é o mecanismo da referência? Em outras palavras, em virtude do quê uma instância de termo referencial conecta-se a um objeto particular/indivíduo? (iv) Existe um único mecanismo de referência comum a todos os termos referenciais, ou os diferentes tipos de termos engatam nos seus referentes em virtude de diferentes tipos de coisas? E (v) em que medida, se é que isto acontece, pode a referência depender de características relativamente privadas do falante⁵, tais como suas intenções ou outros estados mentais?

Nosso objetivo aqui será mapear como as respostas a essas várias questões naturalmente agrupam-se em distintas abordagens da referência, abordagens essas que direcionam grande parte da teorização sobre certos termos ou frases referenciais. Para chegar lá, nós começaremos na seção 2 exami-

⁴ No original, *token uses*. (Nota do tradutor)

⁵ Pronomes pessoais, tais como "ele" e "ela", salvo quando o contexto claramente indicar o contrário, são empregados como termos de gênero-neutro. (Nota do tradutor)

nando as duas principais abordagens dos nomes próprios. Então, na seção 3, nos voltaremos para os indexicais e demonstrativos, o que desafiará o pensamento de que estas duas abordagens representam um conjunto exaustivo de alternativas. A seção 4 se concentrará nas descrições definidas, o que servirá para destacar a questão de como a relação referencial se relaciona com o ato referencial – uma questão que serve para assinalar algumas deficiências potenciais nos modelos anteriores da referência linguística. A seção 5 dará um passo atrás para resumir os principais grupos de posições que desenvolvemos até o momento. Finalmente, a seção 6 encerra examinando as razões para duvidar se realmente existe alguma coisa tal como a relação referencial.

Por razões de clareza e brevidade, este verbete irá abster-se de se aventurar profundamente em debates acerca do significado, forma sintática, ou função dos diversos termos que nós iremos estudar. Ao invés disso, nos concentramos em investigar no que pode ser dito, se é que algo pode ser dito, sobre a referência linguística em geral. Para uma investigação mais detalhada das idiossincrasias destes vários termos, por favor consultar os verbetes sobre nomes, descrições e indexicais. Recomendamos também a consulta aos verbetes sobre tipos naturais.

2. Nomes Próprios

Os nomes próprios há muito tempo ocupam lugar central nos debates sobre a referência linguística. Para os presentes propósitos, os trataremos como aproximadamente co-extensivos com o que os falantes comuns (não-filosoficamente treinados) normalmente denominam de "nomes". Assim, expressões como "Barack Obama", "Kyoto" e "Monte Kilimanjaro" contarão todos como nomes próprios para os nossos propósitos. O que essas expressões têm em comum? Em virtude do quê elas são uma classe genuína de expressões linguísticas? Ao menos à primeira vista, estas parecem ser expressões sintaticamente simples que se referem, ou ao menos pretendem referir, a objetos particulares ou individuais. Assim, "Barack Obama" se refere a um homem particular, "Kyoto" se refere a uma cidade particular e "Monte Kilimanjaro" se refere a uma montanha particular. E, mesmo que seja questionável se expressões como "Papai Noel" e "Sherlock Holmes" realmente se referem a alguma coisa, não há qualquer dúvida que, ao menos, pretendem referir: a Papai Noel e Sherlock Homes, respectivamente. Elas assim contam como nomes próprios para os presentes propósitos.

No que diz respeito aos nomes próprios, existem duas orientações básicas que há muito tempo capturam a atenção dos filósofos: uma que concebe os nomes próprios segundo o modelo das etiquetas, e outra que os concebe segundo o modelo das descrições. Examinaremos brevemente estas duas visões, juntamente com algumas complicações que surgem para cada. Então nos voltaremos para um problema que surge para ambas: explicar como referência funciona para nomes com mais de um portador.

2.1. Teorias Descritivistas

De acordo com as teorias descritivistas dos nomes próprios, um uso particular de nome próprio refere quando o conteúdo descritivo, de alguma forma associado ao uso, seleciona um objeto ou indivíduo. Em uma maneira comum de desenvolver este tipo de concepção, associada a Gottlob Frege (1982) e Bertrand Russell (1911), um conteúdo descritivo particular está ligado a um dado uso de um nome porque o falante associa este conteúdo, em sua mente, ao nome em questão. Para que o uso do nome venha a referir, este conteúdo descritivo precisa determinar um único referente para o nome. Assim, quando o falante usa o nome "N" e, ao fazê-lo, se refere com sucesso a um objeto particular ou indivíduo x, este tipo de descritivista alega (i) que o falante deve pensar em N como o (único) F e (ii) que x deve ser o (único) F. Em outras palavras, este tipo descritivista "clássico" afirma que o sucesso referencial depende do falante associar a cada nome em seu repertório algum conteúdo F que seleciona um único objeto específico. Reciprocamente, quando os falantes não conseguem associar uma descrição suficientemente precisa a um nome, este tipo de descritivismo prevê que a referência falhará.

Descritivistas clássicos, como Frege e Russell, estavam dispostos a reconhecer que o conteúdo descritivo em questão pode variar – algumas vezes de maneira bastante acentuada – de um falante para outro. De fato, segundo Russell, esses conteúdos podem variar ao longo do tempo para um e mesmo falante. Assim, embora eu possa associar o nome "Obama" ao conteúdo descritivo o presidente anterior dos EUA, Michelle Obama pode associar o mesmo nome ao conteúdo descritivo meu marido. Se Trump assumisse poderes ditatoriais, meu conteúdo identificatório associado a Obama poderia mudar – talvez para algo como o último presidente legítimo dos Estados Unidos. Se Barrack e Michelle se divorciassem, o conteúdo identificatório para Michelle sem dúvida também mudaria – talvez para meu ex-marido. Em todos esses casos, o indivíduo referido pelo nome é determinado (ou, como é frequentemente colocado,

"selecionado" ou "fixado") pelo conteúdo descritivo particular que o falante associa ao nome. Porque o conteúdo descritivo em questão é tipicamente caracterizado por meio de uma descrição definida (uma expressão da forma o F), tais teorias são frequentemente (ainda que de maneira um pouco enganadora)⁶ conhecidas como "teorias descritivistas" dos nomes próprios.

Para ter uma visão mais clara do apelo intuitivo do descritivismo, considere um caso no qual conhecemos dois indivíduos nomeados "David Cameron" – um deles é um amigo em comum e nunca exerceu um cargo eletivo, enquanto o outro é o ex-Primeiro Ministro do Reino Unido. Suponha agora que eu assertivamente profira:

2-David Cameron cometeu um grave erro.

Se, por alguma razão, você está confuso sobre quem eu estou falando, é natural que você pergunte "a qual David Cameron você se refere?" Eu posso então responder "ao ex-Primeiro Ministro do Reino Unido", e isto parece estar de acordo com os fatos. Isto é, supondo ser esta a descrição que eu tinha em mente, então parece que o meu proferimento de (2) é genuinamente sobre David Cameron, o ex-Primeiro Ministro, e não sobre nosso amigo em comum, o não-político. De forma semelhante, a verdade ou falsidade de (2) parece depender de como as coisas são com o Primeiro ministro e não de como as coisas são com nosso amigo homônimo.

Alguns outros aspectos do apelo do descritivismo se tornam evidentes quando nós pareamos a tese acima considerada, uma tese acerca do quê determina a referência, com outra tese que naturalmente a acompanha, a saber, a tese de que os conteúdos descritivos associados com os usos dos nomes também fornecem o significado das instâncias dos nomes. Em outros termos, nomes próprios podem muito bem referir, mas talvez eles o façam apenas através dos seus significados, os quais são mais semelhantes às descrições definidas. Ao adotar essa tese, a descritivista tem explicações prontamente disponíveis para um amplo conjunto de casos filosoficamente interessantes. Assim, considere:

3- Héspero é Fósforo

4- Papai Noel vive no Polo Norte.

⁶ Searle (1983) alega que construir o conteúdo determinante das referências como linguisticamente especificável em todos os casos é interpretar mal a natureza de tal conteúdo. Alguns conteúdos podem (por exemplo) ser baseados na percepção, mas não linguisticamente especificáveis. Consultar também Frege (1892).

5- Fred acredita que Cícero, mas não Túlio, era romano.

(3) é verdadeira, mas não cognoscível a priori. Dito isso, se o significado desses dois nomes é apenas a sua referência, nós não temos uma explicação óbvia de por que (3) não é cognoscível a priori. Afinal de contas, a proposição (3) torna-se uma instância da lei da auto-identidade – algo que muito plausivelmente é cognoscível a priori. O descritivista sobre o significado, no entanto, pode dizer que (3) afirma, na verdade, que duas descrições distintas são satisfeitas pelo mesmo objeto. Diante das descrições que seriam relevantes para esse caso, isso não é algo que, plausivelmente, possa ser cognoscível a priori. (4) apresenta um tipo diferente de desafio: se o nome "Papai Noel" falha em referir, então como o proferimento desta sentença pode contar como significativo? Em outras palavras, acerca do quê é esta sentença? Novamente, o descritivista sobre significado tem uma resposta fácil. (4) significa algo como: existe um único elfo barbudo e alegre que vive no Polo Norte. Finalmente, alguém poderia preocupar-se com o fato de que se "Cícero" e "Túlio" apenas referem, e se referem à mesma coisa, então é difícil ver como Fred poderia acreditar algo acerca de um, mas não acreditar o mesmo acerca do outro. Porém, se Fred associa cada nome a uma descrição diferente, e se estas descrições são cruciais para o significado de (5), então o problema simplesmente é dissolvido.

O desafio central à teoria descritivista é que parece haver razão para suspeitar que nomes próprios não são semanticamente equivalentes às descrições definidas. Saul Kripke, por exemplo, argumenta que nomes próprios selecionam o mesmo objeto, ainda quando sob o escopo de termos modais tais como "poderia", ao passo que descrições definidas tipicamente não o fazem. Retomando nosso exemplo anterior de "David Cameron", e supondo que a descrição que eu associo com Cameron é o ex-primeiro ministro que convocou o referendo para o Brexit, parece que eu ainda posso verdadeiramente afirmar:

6-David Cameron poderia não ter convocado o Brexit.

Se o descritivismo está correto sobre o significado e a referência, entretanto, então (6) deve ser equivalente a:

7- Poderia ter sido o caso que: o Primeiro Ministro Britânico que convocou o referendo sobre o Brexit não convocou o referendo sobre o Brexit. Por certo, (7) não é realmente uma alegação do português⁷ comum. Ainda assim, mantendo fixo o fato de que estamos falando de possibilidade metafísica, ao invés de

⁷ No original, *ordinary english*. (Nota do tradutor)

epistêmica, deveria ser suficientemente claro que () é falsa: se há alguém que satisfaz a descrição "o Primeiro Ministro Britânico que convocou um referendo sobre o Brexit" em qualquer situação possível que venhamos a considerar, então esse mesmo indivíduo convocou um referendo para o Brexit nessa mesma situação possível.

Agora, em um português⁸ mais usual, existe uma maneira de entender que considera a frase (6) como verdadeira: assumindo que estamos falando sobre o Primeiro Ministro que atualmente convocou o referendo, é claro que ele poderia não ter convocado. Mas não é assim que tipicamente se presume que as descrições definidas funcionam; ao contrário, normalmente se supõe que elas são flexíveis, ou não-rígidas através do espaço de possibilidades, selecionando o quer que as satisfaça em qualquer situação que venhamos a considerar. Assim, se o descritivista se vale deste tipo de manobra defensiva, como alguns foram tentados a fazer, então ele teria que motivar a alegação de que nomes não são semanticamente equivalentes a descrições definidas per se, mas, ao contrário, às suas versões "atualizadas" ou "rigidificadas"⁹.

Este desafio pode ser evitado por um descritivista disposto a abandonar a alegação que o significado de um nome próprio, em uma dada ocasião, é equivalente a uma descrição definida. Neste caso, uma descrição associada vai fixar a referência relativamente ao mundo atual, e então o referente é o que for relevante para determinar a verdade de enunciados modais. Esta resposta, entretanto, implica abandonar as boas explicações de (3)-(5) que o mais comprometido descritivista do significado estava disposto a fornecer. E ainda mais, mesmo esta versão mínima do descritivismo irá encontrar um outro problema levantado por Kripke, frequentemente denominado de problema "semântico".

Eis o problema: frequentemente, não associamos a um nome informação suficiente para selecionar qualquer indivíduo em particular. Mesmo assim, parecemos capazes de usar este nome para nos referir a um indivíduo específico. Kripke apresenta como exemplo o nome "Feynman". A maioria das pessoas, alega Kripke, irão apenas conhecer Feynman como um físico; elas não saberão nada, além do nome, que serviria para diferenciar Feynman de qualquer outro físico do qual ouviram falar. O problema é que uma descrição indefinida como um

⁸ Vide nota anterior do tradutor. (Nota do tradutor)

⁹ Para uma discussão extensa, consultar Soames (2002). Consultar também o verbete sobre os designadores rígidos.

físico não será suficiente para selecionar qualquer indivíduo particular no mundo. Mesmo um físico denominado "Feynman" tampouco funcionará, ao menos em um mundo no qual dois físicos possuem o mesmo nome. Na melhor das hipóteses, esse tipo de descrição irá selecionar um membro arbitrário de uma classe de indivíduos, não um indivíduo determinado. Além disso, como Kripke aponta, parece perfeitamente coerente para alguém que nada sabe a respeito de Feynman, que apenas ouviu alguém usando o nome, dizer a si mesmo "eu me pergunto quem é Feynman", ou perguntar à seu amigo "Quem é Feynman?". Em cada um destes casos, o natural a dizer é que o falante está usando o nome "Feynman" para perguntar, a si mesmo ou aos outros, a respeito de Feynman. Como ele pode fazê-lo, entretanto, parece muito difícil para o descritivista explicar – supondo (i) que ele quer manter a conexão entre as descrições associadas e a informação disponível ao falante individual, e (ii) que ele não está disposto a confiar em descrições como a quem quer que a pessoa de quem eu ouvi este nome estava usando-o para se referir.

Antes de avançar para a alternativa milliana, é importante notar que alguns descritivistas tentaram defender a teoria abandonando a conexão mencionada em (i). P.F. Strawson (1959), por exemplo, sugeriu que falantes podem confiar nos outros quanto ao fornecimento do conteúdo descritivo relevante, o conteúdo que serve para engatar um dado uso de um nome em um objeto ou indivíduo no mundo. De fato, Strawson permite que grupos efetivamente utilizem nomes para referir, desde que exista entre eles ao menos um conhecedor da referência do nome, para quem o seguinte é verdadeiro: quando coletamos as opiniões dos conhecedores da referência, uma multiplicidade de descrições que eles associam ao nome relevante são verdadeiras de um único objeto. Com isso, o descritivista consegue efetivamente lidar com o caso "Feynman" de Kripke, mas o faz às custas da intuição. Por ora, o descritivista não apenas perdeu a sua capacidade de explicar (3)-(5), como também correu o risco de fazer os falantes não terem qualquer acesso especial à pessoa da qual eles estão falando quando estão usando um nome. Se os falantes regularmente carecem de informação suficiente para identificar o referente do seu próprio uso de um nome próprio, então não fica claro por que estaríamos justificados em confiar neles para responder à questão "A respeito de qual N você está falando?". Alguns podem estar tentados a pagar este preço. Outros podem ser tentados a integrar partes do descritivismo em suas teorias que, fora isso, são teorias anti-descritivistas. De fato, esta última opinião tem se provado popular, e muito do que segue pode ser

visto como um estudo do desempenho desta estratégia com respeito aos diferentes tipos de termos referenciais.

2.2. Herdeiros Millianos

A principal alternativa à teoria descritivista dos nomes é tipicamente denominada de "millianismo". Segundo esta visão, que remonta a John Stuart Mill (1867), o significado de um nome é simplesmente o seu referente. Em sua forma moderna, ela foi introduzida por Ruth Barcan Marcus (1961), que propôs considerar os nomes próprios como "etiquetas". Dizer que o nome próprio é uma etiqueta é, para Marcus, dizer que ele não possui nenhum significado linguístico para além do seu referente. Nomes próprios, de acordo com este tipo de visão, não referem por meio das descrições, as quais eles alegadamente substituem. Ao contrário, eles referem diretamente, como as vezes é afirmado, aos seus portadores. Consequências importantes desta teoria incluem, como assinala Marcus, a consequência de que enunciados de identidade entre termos co-referenciais são verdades necessárias – algo que, ainda que muito intuitivo, não é garantido por muitas teorias descritivistas dos nomes próprios¹⁰. Outras consequências importantes incluem a dissolução de enigmas envolvendo a substitutividade em contextos modais (Marcus 1993).

É claro, dizer que nomes funcionam como etiquetas – ou que eles são "diretamente referenciais" – não é fornecer uma teoria completa acerca dos nomes. Isto vai exigir, ao menos, uma explicação do que faz um nome a etiqueta particular que ele é. Em outras palavras, nós precisamos especificar o que, senão a descrição associada, fixa a referência do nome em um contexto. Colocando de maneira ligeiramente diferente, a alegação de que nomes funcionam como etiquetas efetivamente nos fornece uma semântica dos nomes. O que nos falta é uma metasemântica dos nomes, uma teoria que nos diga qual valor semântico deve ser associado a cada nome, e o porquê.

A opção mais popular foi associar a semântica milliana a uma concepção metasemântica, o que foi sugerido ao mesmo tempo por Peter Geach (1969), Keith Donnellan (1970), e o previamente mencionado Kripke (1972). Tipicamente denominada de "teoria causal da referência", a ideia central desenvolvida nestes trabalhos é que (o uso de) um nome se refere ao que estiver ligado a

¹⁰ Para uma prova formal da necessidade da identidade na lógica modal quantificada, consultar Marcus (1947).

ele da maneira apropriada – uma maneira que não requeira que os falantes associem qualquer conteúdo descritivo identificatório ao nome. A teoria causal é geralmente apresentada com dois componentes: um que lida com a fixação da referência, outro que lida com o empréstimo da referência. A referência, neste tipo de visão, é inicialmente fixada por meio de um ato de nomeação ou batismo. Em outras palavras, o usuário da linguagem dá o nome a um objeto, ao dizer algo como “você será denominado “N””. Supõe-se que no caso padrão aquele que nomeia está presentemente percebendo o objeto no momento em que profere estas palavras. Assim, após o ato inicial de fixação da referência, o nome é transmitido de falante para falante através de trocas comunicativas. Os falantes são bem-sucedidos em referir a algo por meio do nome, segundo esta visão, porque, subjacente aos usos do nome, estão ligações em uma cadeia causal que se estende até o ato inicial de nomeação do objeto. Falantes subsequentes efetivamente “tomam emprestada” a sua referência dos falantes anteriores da cadeia, embora os devedores não precisem ser capazes de identificar qualquer um dos credores dos quais dependem. Tudo o que se requer é que os devedores estejam apropriadamente conectados aos seus credores através de cadeias de comunicação, as cadeias de transferência de nomes.

Como Kripke aponta, complicações surgem devido ao fato de que nós podemos aparentemente reutilizar nomes. Assim, eu posso ter encontrado com o nome “Napoleão” através de uma cadeia de uso que se estende até o mais famoso dos generais franceses. Depois de escutar o nome, eu posso agora decidir nomear o meu gato, o qual gosta de inspecionar imperiosamente o domínio do meu apartamento, de “Napoleão”. Quando eu uso o nome desta maneira, minhas intenções tornam claro, alega Kripke, que eu introduzi um novo nome, ou ao menos o novo uso de um nome. Isto é assim, apesar de existir um sentido muito real no qual eu herdei o nome de uma tradição histórica que remonta à França do século 18. O resultado é este: nem todo tipo de conexão causal a um uso anterior é igual, segundo o teórico causal. Algumas ajudam a fixar a referência, enquanto outras são irrelevantes. Embora seja bom possuir uma teoria completa que permita identificar quais são relevantes e quais não são, foi demon-

strado que é notavelmente difícil fornecer uma explicação adequada para essa distinção¹¹.

2.3. Nomes com Múltiplos Portadores

Até agora, em geral procedemos como se os nomes fossem unívocos – isto é, como se apenas uma pessoa ou coisa pudesse portar um único nome. Porém, isto dificilmente parece correto. Como já vimos, parece que um nome como "David Cameron" pode ser aplicado igualmente bem, tanto ao Primeiro Ministro Britânico como a quaisquer pessoas cujos pais tivessem decidido denominá-las assim, independente da política dos seus respectivos pais; similarmente, uma vez que eu tenha nomeado meu gato de "Napoleão", parece que o nome "Napoleão" pode ser usado igualmente para falar tanto sobre o maior dos generais franceses quanto, alternativamente, sobre meu gato. Descritivistas clássicos não têm qualquer problema com a explicação disto, dado que, a cada uso de um nome, sua referência essencialmente é fixada novamente por qualquer descrição que o falante tenha em mente. Em outras palavras, de acordo com o descritivista, é por mero acaso que os usos dos nomes algumas vezes co-referem. Isto pode parecer um preço muito alto a pagar. Por outro lado, uma vez nos distanciando do descritivismo clássico, explicar como nomes podem referir a diferentes indivíduos em circunstâncias distintas se torna muito mais desafiador.

Considere um descritivista do tipo strawsoniano, para quem o proferimento do nome "N" se refere de acordo com a multiplicidade de crenças dos conhecedores relevantes da referência. O que então faremos com nomes como "David Cameron", o qual às vezes se refere ao ex Primeiro Ministro Britânico, e outras ao nosso amigo comum? Não podemos simplesmente agregar toda a informação relevante associada com esse nome, considerado em sua forma ortográfica ou fonológica, e então verificar quem a pluralidade destas crenças seleciona. Pois isto nos permitiria apenas referir a um destes David Camerons, e isto dificilmente parece correto. Parece, portanto, que uma descritivista deste tipo terá de pensar nos nomes de uma maneira mais refinada. Todavia, isto não será

¹¹ Consultar Evans (1973, 1982) para diversos casos adicionais que se mostraram vexatórios para o teórico causal mais comprometido. Evans também esboça uma importante variação da teoria causal – a qual se concentra na causa da pluralidade de crenças do falante, ao invés da aquisição do nome em si mesmo – e que supostamente resolve a maioria destas preocupações. Para desenvolvimentos posteriores desta linha geral de pensamento, consultar Dickie (2015).

fácil. Ela não pode simplesmente apelar novamente para a informação que o conhecedor da referência tem disponível para a tarefa – pois nós podemos facilmente imaginar alguém que possui nomes para ambos os David Camerons em seu vocabulário e, no entanto, carece de informação suficiente para distinguir um do outro.

Questões semelhantes surgem para o milliano. Aqui, contudo, algumas propostas concretas foram apresentadas. Por exemplo, David Kaplan (1990) argumentou que, contrariamente às aparências, cada nome possui apenas uma referência. De fato, Kaplan sugere que nós devemos pensar os nomes como parcialmente individuados pelos seus referentes. Em outras palavras, de acordo com Kaplan, alguém não pode simplesmente "identificar" qual nome foi utilizado com base na sua forma fonológica ou ortográfica manifesta. Ao contrário, Kaplan afirma que pode existir qualquer número de nomes, todos pronunciados e escritos como "David Cameron". Embora cada um destes nomes seja escrito e pronunciado identicamente, cada um se refere a uma pessoa diferente. A questão então se torna: o que determinada qual nome um falante usou em um dado contexto? Kaplan sugere que a resposta a esta questão tem relação com os estados mentais da falante, e em particular a intenção da falante de falar a respeito deste ou daquele indivíduo. Com efeito, Kaplan considera que escutar o nome produz em nós uma marca mental, análogo mental de um nome linguístico, a qual se refere diretamente a um indivíduo particular. Ao buscar a quem ou a que esta marca se refere, nós passamos a conhecer as propriedades referenciais da enunciação ou inscrição usada para expressar este fenômeno mental. Resumidamente, na teoria de Kaplan os nomes são, em primeiro lugar, um fenômeno mental. Nós, é claro, usamos certos sinais gráficos e sons para externalizar as marcas mentais, ou nomes mentais, que adquirimos uns dos outros no decurso do tempo; nomes linguísticos, no sentido de alguma coisa que possui propriedades referenciais, são tipos de usos de uma forma fonológica e ortográfica particular para externalizar o uso de um nome e para referir em pensamento.

Entretanto, a teoria de Kaplan tem algumas consequências estranhas. Por exemplo, que nunca duas pessoas realmente compartilham o nome – pois um nome vem completo com suas características referenciais. Kaplan sugere que as pessoas não compartilham os nomes, mas, ao contrário, algum aspecto comum de como os seus nomes tendem a ser externalizados. Em outras palavras, segundo Kaplan, compartilha-se um nome com alguém quando ambos os nomes pertencem às mesmas classes fonológicas e ortográficas. Kaplan denomina esses últimos de "nomes genéricos", em oposição aos "nomes

correntes", os quais, para ele, são os portadores de propriedades semânticas como a referência.

Insatisfeitos com as propostas de Kaplan aqui apresentadas, outros teóricos da referência direta adotaram respostas diferentes ao problema dos nomes com múltiplos portadores. Por exemplo, Francois Recanati (1997) e Michael Pelczar e Joe Rainsbury (1998) sugeriram que nomes devem ser tratados segundo o modelo dos indexicais, um tipo de termo que logo discutiremos em mais detalhes. Resumidamente, a sugestão é a seguinte: existe um e único nome "David Cameron", mas o contexto torna claro a qual indivíduo um proferimento particular deste nome se refere. Com efeito, esta proposta preserva algo da teoria causal – ao permitir que batismos e transferências de nomes sejam o tipo de coisa que determina o conjunto de possíveis referentes associados a um nome em um período qualquer – enquanto apela a outra característica do contexto para realizar o trabalho de selecionar um indivíduo deste grupo. Um problema persistente é que, contrariamente aos indexicais como "eu" ou "aqui", não é claro quais aspectos do contexto podem ajudar a fazer esta seleção. As coisas pioram quando passamos a considerar contextos nos quais múltiplas pessoas portadoras do mesmo nome estão todas presentes, e sobre todas elas se fala no decurso de um único proferimento (e.g., "David era Primeiro Ministro, agora afastado após um período de desgraça, enquanto David é um encanador confiável")¹². Uma possibilidade seria afirmar que nomes próprios funcionam mais como os demonstrativos "isto" e "aquilo" do que como indexicais puros. Alternativamente, alguém poderia afirmar que nomes próprios por eles mesmos denotam apenas a propriedade de portar o nome relevante, mas que a referência singular emerge quando os nomes próprios são suplementados com um determinante como "aquele" ou "a" para formar um sintagma determinante, tal como "aquele David" ou "a Michelle" (consultar Burge (1973) e Graaf Fara (2015), respectivamente; consultar também os verbetes sobre nomes para mais sobre tais visões "predicativistas"). Em linguagens como o inglês, por exemplo, estes determinantes acabariam por ser predominantemente silenciosos. Ainda assim, alguém poderia julgar que a existência de determinantes silenciosos é um bom preço a pagar por uma estória coerente sobre a referência dos nomes (ou sintagmas nominais definidos contendo nomes, como pode ser o caso) compatível

¹² Consultar Rami (2014) para discussões úteis sobre estas questões e uma formulação alternativa da visão indexical, desenvolvida para lidar com muitas destas questões.

com a alegação de que dois indivíduos podem genuinamente compartilhar o mesmo nome.

Como já deve estar claro, para melhor entender estas últimas propostas devemos primeiro esclarecer como a referência supostamente funciona para cada um destes diversos termos e frases: indexicais puros, indexicais impuros, demonstrativos complexos e descrições definidas. Nas próximas seções, trabalharemos para preencher estas lacunas.

3. Indexicais

Vimos dois modelos básicos de como as palavras se referem às coisas. No modelo descritivista, palavras referem por estarem associadas, de algum modo, à descrição que isola um objeto particular como o referente. Versões diferentes deste modelo resultam de maneiras diferentes de associar as descrições relevantes aos usos particulares das palavras. No modelo causal, em contraste, palavras são associadas com cadeias de uso que levam de volta a algum ato original de nomeação ou batismo. O ato de nomeação ele mesmo serve, então, para fechar a lacuna entre a palavra e o mundo. Embora ambos os modelos tenham sido desenvolvidos com os nomes em mente, devemos nos questionar: “pode algum deles servir para explicar a referência linguística em geral? Isto é, pode algum dos modelos plausivelmente ser estendido para outros termos referenciais, além dos nomes?”

Para responder a esta questão, e no final introduzir um terceiro modelo distinto de referência linguística, agora nós voltamos para os indexicais. Isto é, nós voltamos para termos como "eu", "você", "aqui", "agora", "ele", "ela", "isto" e "aquilo". Logo veremos, não é nada claro que todos os indexicais referem da mesma maneira. Em particular, uma distinção foi frequentemente feita entre os denominados indexicais "puros" e "impuros", com diferentes teorias da referência sendo propostas para cada. Os desafios que surgem na tentativa de explicar os indexicais puros nos farão retornar a uma das questões com as quais começamos: a saber, as palavras referem porque as usamos para referir, ou nós as usamos para referir porque elas previamente já referem?

3.1. Indexicais Puros

O que são indexicais puros? Grosso modo, eles são expressões referenciais que parecem co-variarem com certos aspectos regulares do contexto nos quais são usados. Aqui, "contexto" deve ser entendido como incorporando, inter

alia, um falante, ouvinte, tempo e lugar. Contrariamente ao que acontece com os indexicais puros, se supõe que a referência de indexicais "impuros" seja mais difícil de caracterizar em termos da seleção de uma propriedade do contexto, distinta e repetível. Embora tanto a existência, como o significado desta distinção sejam controversos (consultar o verbete sobre indexicais para discussões posteriores; consultar também Radulescu (2018)), exemplos de cada um destes tipos tipicamente não são – caso a distinção se prove real e significativa. Exemplos padrões de indexicais impuros incluem "eu", "aqui", "agora" e provavelmente "você". A lista de indexicais impuros, por outro lado, geralmente inclui "este", "aquilo", "ele", "ela" e "isto".

A abordagem tradicional dos indexicais, a qual remonta mais uma vez a Frege e a Russell, afirma que a referência de tais expressões é fixada por algum tipo de conteúdo descritivo associado pelo falante com a expressão. Esta descrição fixadora da referência é o significado de um dado proferimento de uma expressão. A motivação para tal visão é muito intuitiva. Indexicais certamente parecem significar algo, e os seus significados presumivelmente têm algo a ver com a forma pela qual essas expressões referem. Por exemplo, o significado de "eu" é provavelmente o falante deste proferimento e se refere a este indivíduo; o significado de "agora" é provavelmente o tempo deste proferimento e se refere a este tempo. E assim por diante.

Uma objeção óbvia a esta visão é que a referência do termo "eu" não parece sensível a nenhum tipo de conteúdo descritivo que um falante possa vir a associar com o termo. Por exemplo, o fato de que eu poderia associar a descrição o presidente anterior dos E.U.A ao termo "eu" não significa que eu poderia de alguma forma ser bem sucedido em usar o termo "eu" para referir a Barack Obama. Outro problema, discutido longamente em Kaplan (1989b), é que tomar as descrições determinantes da referência como parte do que asserimos quando usamos indexicais pode levar a previsões bastante estranhas. Considere o proferimento assertivo de:

8-Eu estou com fome.

Suponha, primeiro, que eu sou o falante. Eu profiro (8). Agora, suponha que você é o falante. Você profere (8). Embora possa existir um sentido no qual "dizemos a mesma coisa" – isto é, o sentido no qual proferimos a mesma sentença-tipo – existe um sentido no qual nós claramente não dizemos a mesma coisa. Nós também podemos colocar assim: afirmamos coisas muito diferentes. E ainda assim, de acordo com o descritivismo clássico, nós dois afirmamos a

mesma coisa. A saber, afirmamos que o falante está com fome. Tudo o que difere é o contexto no qual nós dois afirmamos isto.

Agora, para deixar claro, existem maneiras de modificar a teoria para evitar ambas as objeções. Alguns, como Hans Reichenbach (1947), ou mais recentemente Manuel Garcia-Carpintero (1998) e John Perry (2001), argumentaram que indexicais são frequentemente "instância-reflexivos", o que significa o seguinte: as descrições associadas aos termos deverão envolver referência explícita à própria instância de uso do termo. Assim, por exemplo, a descrição de "eu" poderia ser: o falante desta própria instância de "eu". Uma vez que diferentes instâncias estarão envolvidas quando cada um de nós profere (8), nós não mais estaremos afirmando a mesma coisa. Se nós acrescentarmos que, para instâncias indexicais ao menos, as descrições relevantes não estão sob o controle do falante, porém estão associados com termos particulares através das regras da linguagem, então nós podemos evitar as objeções anteriores também.

Um tipo diferente de abordagem foi desenvolvida pelo aluno de Reichenbach, David Kaplan. Kaplan (1989b) tomou muitos dos mesmos elementos com os quais Reichenbach trabalhava, mas os inseriu em uma abordagem de referência direta, na qual o significado de um indexical em um contexto – a sua contribuição para o que é dito ou afirmado – é apenas um objeto. De acordo com Kaplan, precisamos distinguir entre dois tipos de significado, os quais denominou de "caráter" e "conteúdo". Conteúdo é basicamente o que até aqui denominamos significado; é a contribuição do proferimento de um termo individual para o que é dito ou afirmado pelo proferimento de uma sentença completa da qual é parte. De acordo com as teorias da referência direta, o conteúdo é um objeto, o que, por conseguinte, torna a verdade ou falsidade da afirmação dependente do objeto. Caráter, por outro lado, é mais semelhante a uma regra de uso; um caráter nos diz, para um contexto qualquer, qual é o conteúdo de uma dada expressão. Os nomes, como Kaplan pensa a respeito deles, possuem caracteres constantes: em dois contextos arbitrariamente escolhidos, usos do mesmo nome serão mapeados no mesmo referente (lembre que Kaplan pensa em nomes a partir uma análise mais refinada, na qual eles nunca são partilhados). A regra para qualquer nome "N" assim se torna: em qualquer contexto, retorne N como o referente. O mesmo não pode ser dito para termos como "eu" e "aqui", os quais Kaplan associa com regras como retorne o falante e retorne a localização da qual parte o proferimento. Assim, em um contexto no qual eu estou falando no bar do Foyle, proferimentos destes indexicais irão referir a mim e ao bar do

Foyle, respectivamente. Quando você está falando no American Bar, eles irão referir a você e ao American Bar.

É importante ressaltar que os caracteres dos indexicais puros não são sensíveis aos estados mentais dos falantes. Isto exclui qualquer possibilidade que eu venha a usar "eu" para me referir a Barack Obama. Tampouco são o meu e o seu proferimentos de (7) predicados ou afirmados da mesma coisa. Além disso, conforme aponta Kaplan, esta visão nos permite produtivamente distinguir entre necessidade "metafísica" e o que ele denomina de necessidade "lógica". A sentença "eu estou aqui agora", alega Kaplan, representa a necessidade lógica: em virtude do significado dos indexicais "eu", "aqui", e "agora", esta sentença não pode ser proferida falsamente. Todavia, claramente não é necessário, em um sentido metafísico qualquer, que uma falante particular esteja no lugar em que ela está no momento do proferimento; ela poderia igual e facilmente estar em algum outro lugar qualquer¹³.

Agora temos à nossa disposição três modelos básicos da referência, da metasemântica dos termos referenciais: o modelo descritivo, o modelo da cadeia causal, e o modelo do caráter. Com o último desses em vista, podemos esclarecer um aspecto importante da teoria indexical dos nomes que foi discutida no fim da última seção. Essencialmente, esta teoria propõe misturar os nossos segundos e terceiros modelos. No lugar de nomes que contêm caracteres constantes, como Kaplan afirmou, a proposta atribui a eles alguns caracteres mais interessantes. Por exemplo, nós podemos tentar estipular que o contexto, além de conter um falante, lugar, tempo, etc., também deve incluir o mais saliente indivíduo que porta o nome dado. Nomes, de acordo com esta ideia, sempre referem ao mais saliente portador em um dado contexto. Uma vez que nomes não mais possuem caracteres constantes, eles são mais como indexicais do que inicialmente pensávamos. Por outro lado, podemos preservar uma boa parte da teoria causal,

¹³ Uma literatura significativa floresceu a respeito desta alegação, centrada no que foi denominado "o paradoxo da máquina que responde". Para uma revisão recente, consultar Cohen e Michaelson (2013). Consultar também o verbete sobre indexicais para discussões e referências adicionais.

ao recharacterizá-la como uma teoria acerca do que conta como portador de um nome em um dado contexto¹⁴.

Esta maneira de desenvolver a teoria indexical, todavia, não está isenta de características negativas. Por exemplo, dificilmente é verdade que nós sempre usamos nomes para referir ao mais saliente portador do nome em um dado contexto, ainda mais quando passamos a considerar proferimentos de nomes prefixados por sintagmas como "o outro" (p.ex, "O outro David não está aqui ainda, mas eu posso ouvi-lo a partir do corredor"). E ainda mais: existe razão para suspeitar que este tipo de teoria indexical implica o abandono de uma das alegadas vantagens da teoria causal: sua habilidade em explicar como conseguimos transmitir a capacidade de pensar sobre objetos meramente com base na transmissão de nomes dos objetos de um falante para o outro. Uma vez que nomes estão causalmente ancorados aos objetos que nomeiam, segundo a teoria causal, adquirir um novo nome deveria ser suficiente para nos colocar em contato causal com o objeto nomeado. De acordo com a teoria indexical dos nomes, por outro lado, não transmitimos os nomes das coisas; ao contrário, nós os usamos para referir ao portador mais saliente do nome em um dado contexto. Talvez isto seja suficiente para transmitir o uso de um nome – carregado, por assim dizer, com um referente – quando o ouvinte não possui previamente o uso em seu repertório. Porém, se o ouvinte não acrescenta este uso ao seu repertório, então se torna difícil ver como o referente deste uso poderia contar como o portador mais saliente do nome relevante neste contexto.

3.2. Indexicais Impuros

Em contraste com os indexicais impuros, a referência dos indexicais impuros como "ele", "ela", e "isso" ou os "verdadeiros demonstrativos" como "isto" e "aquilo" parece menos sujeita a uma análise em termos do caráter. Afinal de contas, qual é a característica regular do contexto discursivo que qualquer um destes termos poderia servir para selecionar? Uma opção seria dizer que instâncias de cada um destes termos simplesmente se referem ao demonstratum no contexto. Porém, isto não apenas ameaça nos forçar a dizer que todos estes termos efetivamente significam a mesma coisa – o que parece muito contraintuitivo

¹⁴ No entanto, consultar Gray (2014, 2015) e Jeshion (2015) para uma razão para pensar que tal explicação do portador do nome não será plenamente adequada. Também é digno de nota que a necessidade de desenvolver uma explicação do portador do nome é um desafio particularmente saliente para os predicativistas sobre nomes próprios.

– como agora também precisamos saber o que exatamente serve para fazer com que algo seja o demonstratum em um determinado contexto. E o que faremos quando diversos termos como estes, ou múltiplas instâncias do mesmo termo, forem usados no período de um mesmo proferimento? Iremos realmente acreditar que envolvem não apenas a função do demonstratum, mas também a função do primeiro demonstratum, a função do segundo demonstratum, etc.?

Várias respostas possíveis surgem neste momento. Primeiro, podemos tentar tirar vantagem do fato de que muitos destes termos são acompanhados por gestos ostensivos. Talvez, conforme sugerem Kaplan (1978) e McGinn (1981), sejam tais gestos que transformam tais objetos nos demonstrata dos respectivos contextos – um por cada uso de um indexical impuro. Esta sugestão também proporciona uma maneira natural de entender o que faz algo o primeiro demonstratum em oposição ao segundo, e assim por diante. Por outro lado, esta abordagem enfrenta uma objeção bastante séria: muitos usos perfeitamente adequados de indexicais impuros e demonstrativos simplesmente não são acompanhados por qualquer tipo de gesto ostensivo. Esta observação leva a uma segunda resposta possível, a qual é, de muitas maneiras, uma generalização da primeira: talvez não seja o gesto ostensivo em si mesmo que faz um objeto o demonstratum, mas aquilo que próprio gesto produz. Gestos, nós podemos dizer, servem para salientar os objetos em um contexto. Então, talvez o importante seja que um objeto é particularmente saliente em um contexto; talvez isto faça um demonstratum, seja esta saliência o resultado direto de um gesto ostensivo ou não.

Howard Wettstein (1984) desenvolve uma visão nesse sentido, e esse mesmo tipo de visão foi mais recentemente elaborada e defendida por Allyson Mount (2008). A ideia básica é que um termo como "ela" se refere a quem contar como a mulher mais saliente no contexto. Regras semelhantes se aplicarão aos outros indexicais impuros. Problemas começam a surgir, entretanto, quando consideramos usos simples de "isto" ou "aquilo" – os quais não parecem conter qualquer informação substancial sobre o termo *sortal*¹⁵ relevante. Embora possa parecer suficientemente plausível que algum objeto contará como a mulher max-

¹⁵ No original, *sortal*. Um termo *sortal* caracteriza-se por ser contável. Termos não *sortais* são termos tais como "água" ou "leite", pois não é comum falarmos "um leite" ou "uma água". Existe uma extensa discussão na linguística a respeito da aplicação desta distinção ao português, e se existem contextos nos quais termos como água ou leite poderiam ser utilizados como *sortais*. (Nota do tradutor)

imamente saliente no contexto, é muito menos plausível que contextos conterão algum objeto como maximamente saliente simpliciter, ou sem qualificação adicional. O que poderia fazer um objeto maximamente saliente simpliciter? A sugestão natural seria: nossos interesses. Se algo fará um objeto maximamente simpliciter, será presumivelmente o fato de que nós, enquanto agentes, trouxemos certo conjunto de interesses à troca conversacional.

Dialeticamente, entretanto, esta sugestão é problemática. Por um lado, os interesses do falante e do ouvinte podem facilmente divergir. Assim, precisamos conhecer qual interesse triunfa sobre o outro, se é que algum deles triunfa, e quando eles não se alinham. Por outro lado, tudo o que podemos dizer é que certo objeto é maximamente saliente em relação tanto ao contexto, quanto ao agente particular – mas nunca relativamente ao contexto em si mesmo, sem qualificações adicionais. Mount especifica que estamos interessados na sua saliência mútua máxima. Porém, isto ameaça fazer com que muitos usos de indexicais impuros não refiram, quando os interesses do falante e do ouvinte divergem. Além disso, conforme Mount aponta, a visão implica que a referência falha em quase qualquer circunstância na qual um ouvinte não esteja prestando atenção ao proferimento do falante. Outras maneiras de tornar a teoria mais precisa enfrentam dificuldades análogas, aparentemente dando muito controle ao falante ou ao ouvinte.

Isto nos traz uma terceira possibilidade acerca de como os indexicais impuros referem: talvez, como Kaplan (1989a) sugere, os usos-instâncias de indexicais impuros, como os demonstrativos e usos dêiticos de pronomes, refiram a quem ou a quê o falante tenha intenção de fazê-los referir. Kaplan vê aqui alguma continuidade com a sua abordagem anterior baseada no apontar, uma vez que ele considera que tais intenções (as quais ele denomina "intenções direcionais") poderiam muito bem ser pensadas como um análogo interno de um apontar externo. Ainda assim, a sugestão é mais duradoura que a anterior; nenhum gesto é necessário para a referência ser bem-sucedida de fato. Com efeito, não apenas podemos agora explicar o uso de indexicais impuros sem qualquer gesto ostensivo, como esse tipo de teoria intencionalista também pode permitir o uso de indexicais impuros para se referir a objetos fisicamente ausentes no contexto – visto não haver impedimento para os falantes terem intenções direcionadas a tais objetos. Uma vez que nós de fato usamos indexicais impuros para referir a objetos físicos e temporalmente distantes, isso parece uma vantagem significativa sobre a teoria "intencionalista" da referência de Kaplan.

Nada disso quer dizer que a teoria de Kaplan foi universalmente aceita. Muito pelo contrário, frequentemente se pensou que a teoria enfrenta um problema bastante sério. Notamos acima que os demonstrativos simples "isto" e "aquilo" parecem bastante flexíveis em sua aplicação. Todavia, mesmo estes termos não são infinitamente flexíveis. Presumidamente, você não pode apontar direta e deliberadamente para uma imagem de David Cameron, proferir (9), e ser bem-sucedido em referir à imagem de David Kaplan que você escondeu na gaveta da sua escrivaninha justamente para tais ocasiões:

9-Esta é a imagem de um dos filósofos do século XX que mais gosta de camisas floridas.

Este é um exemplo do que às vezes foi denominado de "problema Humpty Dumpty"¹⁶ Basicamente, sempre que nós formos tentados a sugerir que o significado ou a referência da instância de uso de algum termo particular depende das intenções do falante, o seguinte tipo de problema inevitavelmente surge: a menos que imponhamos algumas restrições sobre estas intenções, então obteremos resultados aparentemente absurdos, tais como o resultado de que a instância de uso de "isto" em (9) realmente se refere à imagem de Kaplan que você escondeu na gaveta da sua escrivaninha. Um caso levemente mais complicado, surpreendente proposto anteriormente pelo próprio Kaplan (i.e. Kaplan, 1978), e utilizado para motivar a teoria do apontar, pode ser utilizado para demonstrar o mesmo ponto. Kaplan pede para imaginarmos um cenário no qual a falante está sentada em sua mesa, e a sua prezada imagem de Rudolf Carnap, normalmente pendurada na parede atrás dela, foi trocada pela imagem de Spiro Agnew. Não notando a troca, e pretendendo usar o termo "isto" para referir à imagem de Carnap, a falante aponta para trás de si, diretamente para a imagem de Agnew, e profere (10):

10-Esta é a imagem de um dos maiores filósofos do século XX.

Na maneira como Kaplan escreve, (10) não é falsa em virtude de uma falha referencial. Ao contrário, é falsa em virtude da falante ter afirmado alguma coisa falsa acerca da imagem atualmente pendurada, a imagem de Agnew. A maioria dos filósofos se inclinou a concordar com isto (entretanto, consultar King (2013) para uma exceção). É importante ressaltar que este caso é levemente

¹⁶ A terminologia foi introduzida na filosofia da linguagem por MacKay (1968). Trata-se de uma alusão a alguns dos comentários memoráveis de Humpty Dumpty em *Alice através do espelho* de Lewis Carroll, os quais se referem ao grau de controle que ele considera ter sobre o significado das palavras que profere.

diferente do nosso (9), pois não parece totalmente implausível que ela pretenda também utilizar o "esta" para se referir à imagem pendurada atrás dela, qualquer que seja a imagem. Afinal, ela está de alguma forma confusa sobre qual imagem é esta. Esta confusão pode, portanto, dar ao intencionalista um espaço adicional de manobra.

Uma possibilidade para o intencionalista seria contestar os julgamentos acima e aceitar a alegação de que as instâncias de uso de "isto" em (9) e (10) de fato se referem às imagens de Kaplan e Carnap, respectivamente (consultar Radulescu (no prelo) para uma resposta nesse sentido). No entanto, a maioria dos intencionalistas não seguiu esse caminho. Em vez disso, eles tentaram restringir a gama de intenções que contam como apropriadamente "referenciais", ou tentaram impor certas limitações sobre quando as intenções dos falantes de fato determinam a referência. Esta primeira estratégia geralmente procedeu recorrendo a certas considerações extraídas da teoria do significado de Paul Grice (1957). A segunda, ao contrário, tentou derivar as restrições relevantes a partir dos significados permanentes ou das regras de uso dos indexicais impuros e demonstrativos. Mais recentemente, alguns filósofos também tentaram combinar estas duas abordagens.

A abordagem griceana da referência foi pela primeira vez claramente articulada por Gail Stine (1978) – embora tenha sido mencionada pelo próprio Grice e por Keith Donnellan (1968)¹⁷. A ideia básica é que os tipos de intenções relevantes para fixar a referência dos indexicais impuros são, propriamente falando, intenções destinadas a chamar a atenção do ouvinte para identificar um objeto particular como o referente. Acompanhando a teoria mais ampla do significado de Grice, afirma-se que essas intenções são ainda mais complicadas do que isso: elas não apenas visam fazer o ouvinte identificar algum objeto como referente, mas também devem ser cumpridas, ao menos em parte, com base no reconhecimento, por parte do ouvinte, de que é assim que ele (o ouvinte) deve agir.

Após ter seguido a linha griceana até aqui, precisamos agora nos perguntar: pode a tese griceana acerca da complexidade das intenções referenciais de alguma forma eliminar a ameaça de um colapso em direção a uma postura estilo Humpty Dumpty? A resposta, segundo o griceano, é "sim". Mais precisamente, o griceano alegará que a falante carece de genuína intenção referencial

¹⁷ Este tipo de visão foi também desenvolvido e defendida por, entre outros, Schiffer (1981) e Bach (1992).

em um caso como (9), pois ele não deveria ter a expectativa de que o seu ouvinte estará em posição de recuperar o referente visado por ele (o falante). Em um caso como (10), entretanto, ele parece ter intenção de fazer a sua ouvinte recuperar a imagem de Carnap como o referente. Então como o griceano está em qualquer condição de ajudar aqui? Como Bach (1992) prestativamente esclarece, a abordagem griceana na verdade deveria funcionar da seguinte maneira: embora genuinamente referenciais, a intenção do falante de se referir à imagem de Carnap é realizada através da intenção de se referir à imagem atrás dela, qualquer que seja a imagem. Essa última intenção deve ser reconhecida diretamente, enquanto que a primeira pode apenas ser reconhecida indiretamente, mediante reconhecimento da segunda intenção, e (nos casos bons) a partir do fato de que a imagem atrás do falante realmente é a imagem de Carnap. Então, uma maneira melhor de formular a tese griceana sobre a referência é em termos de uma distinção primária/secundária, a qual reflete este tipo de assimetria: quando existe um conflito entre as intenções referenciais primárias e secundárias, e ambas são genuinamente referenciais, é a primeira que fixa a referência¹⁸.

Até aqui, tudo ótimo para o griceano. Porém, quão geral é a solução para o problema de Humpty Dumpty? Pode ele, por exemplo, explicar um falante que parece plenamente competente em uma linguagem como português, mas que pensa falsamente que todos podem simplesmente adivinhar as suas intenções diretamente? Em outras palavras, considerem um falante que acredita que o seu ouvinte pode ler a sua mente, porém apenas no que diz respeito às suas intenções referenciais, mas nada além disso, e apenas quando ele profere um demonstrativo. A menos que excluamos tal falante como amplamente incompetente em falar português, parece que ele pode ter sucesso, segundo o griceano, em fazer as instâncias de "isto" e "aquilo" se referirem ao quê ele quiser, em qualquer ocasião. Pois este falante pode pretender coerentemente que seu ouvinte compreenda esse objeto como o referente do seu uso do demonstrativo, parcialmente em virtude do reconhecimento desta intenção primária ela mesmo, e independentemente de qual objeto for. Em outras palavras, supondo algumas crenças suficientemente estranhas, o griceano prevê que o falante realmente pode fazer o uso de "isso" em um proferimento como (9) para se

¹⁸ Consultar também King (2013) por uma solução surpreendentemente similar. Para uma discussão útil e críticas, consultar Speaks (2016, 2017), Unnsteinsson (no prelo), e Radulescu (no prelo).

referir à sua imagem de David Kaplan – independentemente de alguém que não seja um clarividente real ser capaz de reconhecer a referência. Assim, a estratégia griceana para bloquear o colapso do intencionalismo em uma postura do tipo Humpty Dumpty não seria completamente generalizada.

Parcialmente em resposta a problemas como esses, podemos pensar que o tipo de resposta ao Problema de Humpty Dumpty não deve ser aumentar a complexidade das relações referenciais em si mesmas, mas, ao contrário, limitar como certos tipos de termos – inclusive os verdadeiros demonstrativos – podem ser usados para referir. Marga Reimer (1991, 1992) oferece uma sugestão neste sentido: quando usos dos demonstrativos "isto" e "aquilo" são acompanhados por gestos ostensivos, então o referente deve estar na direção geral indicada pelo gesto¹⁹. Quando não há gesto, então não existe tal restrição. Em outras palavras, Reimer enfatiza novamente a conexão aparentemente especial entre demonstrativos e gestos que os teóricos anteriores tentaram transformar em uma teoria completa da referência demonstrativa. Desta vez, entretanto, Reimer sugere que nós não usemos a conexão para gerar uma teoria completa da referência demonstrativa, mas, ao contrário, para impor algumas restrições sobre a teoria intencionalista. Estas restrições, por sua vez, permitem a Reimer fazer previsões intuitivamente corretas sobre (9) e (10): com respeito a (9), irá dizer que não se pode ser bem sucedido em apontar para uma imagem e se referir a outra, a menos que talvez a primeira seja uma fotografia da segunda; quanto a (10), dirá que a falante pretende se referir a ambos, tanto à imagem atrás de si quanto a Carnap, mas apenas uma dessas satisfaz a restrição associada ao gesto. Assim, apenas uma destas intenções está em posição de determinar a referência.

Para esclarecer, a posição de Reimer não é, de modo algum, incompatível com a teoria griceana das intenções referenciais. Pode-se, de fato, aceitar que existem restrições substantivas sobre a referência derivadas dos significados dos indexicais impuros, e também que intenções referenciais são como os griceanos descrevem-nas. Bach (2017) recentemente aceitou a tese

¹⁹ Um tipo diferente de restrição sobre a referência demonstrativa foi desenvolvido em Neale (2004), Stokke (2010), e King (2014). Prescindindo das várias diferenças entre estas explicações, cada qual sustenta que o sucesso referencial requer que o referente seja recuperável ou atualmente recuperado pelo ouvinte (ou por uma versão idealizada do ouvinte). Entretanto, pouco detalhes foram fornecidos a respeito dos tipos de idealizações previstas, que torna certas versões deste tipo de visão um tanto difíceis de avaliar.

griceana, por exemplo. Pois parece que pode-se prevenir o colapso no problema de Humpty Dumpty simplesmente aceitando a existência de restrições sobre a referência convencionalmente associadas aos indexicais impuros; assim não seria mais necessário que a tese griceana tenha aplicação no escopo das intenções referenciais. Ainda assim, é perfeitamente razoável pensar que existem muito mais razões teóricas gerais para preferir a combinação destas duas teorias em detrimento de uma abordagem centrada na restrição, tal como a de Reimer.

4. Descrições Definidas

Em contraste com os tipos de termos que consideramos até agora, a questão fundamental a respeito das descrições definidas – expressões complexas como "o biscoito" e "a loja de conveniência" – não é como elas referem, mas se elas referem.

Meinong (1904) e Frege (1892) pensavam que sim, essas expressões referem. Isto gerou um problema para eles, entretanto, pois o material descritivo manifesto encontrado na descrição definida é tipicamente insuficiente para identificar um único objeto como o referente. Para Frege, ao menos, existia uma solução bastante óbvia: tratar o material descritivo manifesto como apenas uma parte do sentido da descrição definida, permitindo a adição de mais material por qualquer descrição mais completa que o falante venha a ter em mente. Confrontado com este mesmo conjunto de questões, Russell (1905) tomou a direção oposta: ele postulou que as descrições simplesmente não referem. Em vez disso, ele ofereceu um procedimento de tradução da forma "o A é B", cujo significado ele postulou como o seguinte: existe um e apenas um A, e este A é também B²⁰.

Esta disputa se torna mais interessante quando se considera que Russell e Frege concordavam que nomes devem ser entendidos, em algum sentido, como similares às descrições definidas. Agora estamos em uma posição melhor para ver as diferenças entre eles, apesar desta concordância superficial. Para

²⁰ Outro desafio, que a concepção russelliana tentou evitar, tem relação com descrições que, intuitivamente, se referem a objetos não existentes. Descrições como "o filho de Deus" ou "o protagonista da trilogia *Broken Earth*" não parecem referir a objetos reais, mas dificilmente são sem sentido. O russelliano pode facilmente acomodar isto mediante o procedimento da tradução, enquanto o referencialista terá que contar uma história adicional a respeito de como estas contam como significativas. Note que um problema análogo, o problema dos nomes vazios, também surge para os millianos. Para mais sobre estas questões, consultar o verbete sobre objetos não existentes.

Frege, essa alegação de forma alguma impedia que os nomes fossem dispositivos genuínos de referência. Para Russell, impedia. Para Frege, a contribuição de um nome para a verdade ou falsidade de uma sentença era (contextos intencionais à parte) o seu referente. Para Russell, era a descrição associada.

Por que supor que descrições definidas não referem? Afinal de contas, nós certamente parecemos usá-las para falar sobre objetos particulares; quando eu digo “eu quero o barco à direita daquele outro”, eu pareço expressar o desejo de possuir um barco determinado, não expressando, de alguma forma complicada, o desejo de ter algum barco qualquer. Além disso, podemos substituir descrições definidas por nomes e indexicais na maior parte dos contextos linguísticos. Pode parecer razoável, ainda que seja ingênua, a afirmação que se nomes e indexicais referem, então as instâncias de descrições definidas também o fazem. Não faltou a Russell, entretanto, motivação para sua alegação de que descrições definidas não referem. Considere proferimentos assertivos das seguintes sentenças:

11-O rei da França é calvo.

12-O autor de *Middlemarch* é o terceiro filho de Robert e Christiana Evans.

(11) possui significado, e certamente não é verdadeira. Russell considerava que ela é evidentemente falsa, embora isso seja controverso (consultar Strawson (1950) e, mais recentemente, Von Fintel (2004)). Independentemente disso, se alguém pensa que descrições definidas são dispositivos de referência, então não está claro como alguém poderia estar justificado ao afirmar que frases como (10) são significativas. Afinal de contas, a descrição claramente não refere. Quanto a (12), o problema que o referencialista enfrenta é diferente: uma vez que ambas as descrições deveriam referir ao mesmo indivíduo, Mary Ann Evans (ou “George Eliot”), é obscuro o motivo pelo qual (12) não deveria ser cognoscível a priori. Assim como acontece com enunciados de identidade envolvendo nomes co-referenciais, esta parece ser apenas uma exemplificação da lei da auto-identidade.

Note-se que a teoria de Russell, em contraste com teorias que postulariam que as descrições definidas referem, não tem qualquer dificuldade em explicar tudo isto. (11) apenas faz uma alegação sobre existir um e apenas um rei da França, uma alegação que é significativa e claramente falsa. (12), por outro lado, faz uma alegação sobre o único indivíduo a satisfazer um certo conjunto de

propriedades, e afirma que este também é o único a satisfazer outro conjunto de propriedades. Isto dificilmente deveria contar como algo cognoscível a priori.

Mesmo assim, ainda existem razões para questionar a alegação de Russell de que descrições definidas não referem. Por exemplo, como Strawson (1950) apontou, muitos proferimentos assertivos envolvendo descrições definidas parecem verdadeiros mesmo quando a descrição definida não é satisfeita por um único objeto. Assim, considere:

13-A mesa está coberta por livros.

Proferida assertivamente em um contexto no qual existe uma única mesa com muitos livros empilhados, (13) pareceria verdadeira. Todavia, Russell previu que o oposto aconteceria, desde que exista ao menos uma mesa adicional em algum lugar do universo. A alternativa de Strawson foi alegar que existem usos referenciais genuínos das descrições definidas. Usadas desta maneira, descrições definidas, assim como os nomes e os indexicais, contribuirão com um objeto ou indivíduo para o conteúdo asserido pela sentença – desde que este objeto satisfaça o material descritivo relevante. Caso isso não aconteça, tais instâncias ainda contarão como significativas em virtude de estarem associadas a regras uso coerentes, mas não conseguirão referir. Isso, Strawson alega, significa que as sentenças assertivas das quais são uma parte falharão em ser verdadeiras ou falsas. A categoria contrastante aqui é a categoria de uso atributivo de uma descrição definida, a qual supõe-se funcionar mais ou menos como Russell sugeriu.

Keith Donnellan (1966) foi ainda mais longe do que Strawson, alegando que descrições definidas podem ser usadas para referir, mesmo quando o presumido referente não satisfaz o material descritivo explícito na descrição. No seu provavelmente mais famoso exemplo, Donnellan pede que consideremos o proferimento seguinte:

14-Quem é o homem segurando o martini?

Suponha que, quando você profere (14), nós dois estamos olhando para o homem que parece estar bebendo o martini, mas que ele está de fato bebendo água. Da maneira como Donnellan descreve, isso é irrelevante; a descrição-instância "o homem bebendo o martini" pode referir ao homem em nosso campo visual, independentemente de ele estar de fato bebendo martini. O que importa, de acordo com Donnellan, é que você tenha este homem "em mente" quando usa a descrição relevante, que você está usando a descrição com este homem como alvo.

Alguém poderia suspeitar, como Alfred MacKay (1968), que isto significa que Donnellan está vulnerável à acusação de fornecer uma visão de tipo Humpty Dumpty para as descrições definidas – na qual não há restrições substanciais sobre as possibilidades de referência das instâncias de descrições. Curiosamente, Donnellan (1968) respondeu apelando para teoria griceana do significado, levantado assim a possibilidade de que o que ele queria dizer com "ter em mente" era algo similar à noção de Grice de ter uma intenção referencial. Como vimos acima, entretanto, existe razão para pensar que mesmo este tipo de manobra não será suficiente para previnir, de uma maneira completamente geral, o colapso do intencionalismo no problema de Humpty Dumpty. Portanto, ainda existe razão para temer que a disposição de Donnellan, de abandonar o material descritivo explícito como uma restrição ao sucesso referencial, talvez vá longe demais.

E a respeito das intuições por trás do exemplo de Donnellan? Pois, mesmo se nós rejeitarmos a sua alegação de que uma descrição como "o F" pode ser usada para referir a algo que não é um F, ainda assim parece como se a comunicação fosse possível em um caso como (14). Isto é, eu deveria ser capaz de saber sobre quem você está falando, independentemente de eu saber que este homem está bebendo água, ao invés de martini. Kripke (1977) sugeriu que a maneira de lidar com casos como estes é enxergar o tipo de referência envolvida como meramente pragmática, tendo eficácia comunicativa ao invés do tipo estrito de direcionalidade ao objeto²¹ relevante para determinar a verdade ou falsidade em um contexto. De acordo com Kripke, não existe necessidade de postular que descrições definidas alguma vez referem; muito pelo contrário, nós podemos dissolver as suas aparentes propriedades referenciais notando que, mesmo se elas não referirem, falantes tipicamente terão alvos em sua mente aos quais recorrer. Enquanto formos razoavelmente bons em recuperar estes alvos, podemos usá-los para direcionar os pensamentos uns dos outros a objetos particulares, sem necessidade de postular que instâncias de uso destas expressões conseguem se conectar a estes objetos e indivíduos. Em outras palavras, somos livres para aceitar a teoria russelliana das descrições, suplementada por esta alegação de que descrições são frequentemente usadas para direcionar a atenção do ouvinte ao "referente do falante", como Kripke denominou, o qual ele julgava ser o análogo referencial do "significado do falante" griceano (ou, grosso modo, o que falante está tentando transmitir ao ouvinte, literalmente ou não).

²¹ No original, *aboutness*. (N. do T.)

Kripke parece pensar que a referência linguística – ou o que denomina "referência semântica" – precisa funcionar exclusivamente por meio de uma convenção. Os seus exemplos paradigmáticos são os nomes, os quais ele toma como associados com uma convenção, no sentido de que as suas referências são fixadas através de um apelo às cadeias causais. Alternativamente, ele poderia ter apontado para a abordagem baseada no caráter para os indexicais impuros, defendida por Kaplan, uma visão que depende igualmente de convenções para estabelecer a referência. Como vimos acima, entretanto, plausivelmente existem limites a esse tipo de abordagem: uma abordagem convencionalista à referência dos indexicais impuros, tais como demonstrativos e pronomes, não se mostrou tão promissora. De forma similar, quando passamos a levar a sério o problema dos nomes com múltiplos portadores, menos claro se torna que algo tão simples como uma convenção linguística pode fixar a referência dos nomes. Pelo contrário, em ambos os casos, parece que será tentador apelar para fatos sobre o falante para fixar a referência linguística. Porém, se isso é aceitável, a defesa das descrições definidas como não referenciais começa a parecer muito mais fraca.

Se, diferentemente de Kripke, nós começamos com a ideia de que a referência linguística é, na sua raiz, o resultado de um ato, de algo que nós fazemos, ao invés de algo que as palavras elas mesmas fazem, então descrições definidas começam a parecer mais como os casos normais e os indexicais como os casos desviantes. Esta é, de fato, uma ideia antiga, que remonta a Susan Stebbing (1943) e ao trabalho do último Ludwig Wittgenstein (1958), e que já vimos presente nas ideias, ideias de Grice, Strawson e Donnellan²². A preocupação permanente, é claro, é que, sem algumas restrições sobre o sucesso referencial dos termos, divorciaremos a teoria da referência das nossas intuições e de qualquer função em uma teoria geral da comunicação. Porém, se aceitarmos certos limites sobre o sucesso referencial – a serem impostos pela reflexividade griceana, restrições convencionadas ou ambas – estas preocupações plausivelmente podem ser evitadas. E, de fato, no que diz respeito às descrições definidas, a argumentação em defesa de restrições convencionais sobre o sucesso referencial, como Strawson acreditava haver, pode parecer particularmente atraente.

²² Consultar Devitt (1997) e Reimer (1998) para uma aplicação destas ideias ao presente contexto. Para críticas, consultar Schoubye (2011).

Mais recentemente, um tipo diferente de desafio emergiu para quem alega que instâncias de descrições definidas algumas vezes referem: a saber, a diferença em significado que normalmente supõe-se distinguir as descrições dos seus similares indefinidos (isto é, descrições como "um filósofo" em oposição a "o filósofo") foi ela mesma questionada. Duas razões principais foram fornecidas para rejeitar a diferença de significado entre estes dois tipos de frases. Primeiro, determinar exatamente como esta distinção deve ser feita tem se mostrado uma tarefa notavelmente difícil; cada aspecto da análise original de Russell, por exemplo, foi subseqüentemente colocada em dúvida. Segundo, muitas linguagens não possuem algo como a distinção definido/indefinido. Isto levanta a possibilidade que esta distinção é uma característica do significado muito menos significativa, e menos bem definida, mesmo em uma linguagem como o inglês, do que normalmente é suposto ser o caso (consultar Ludlow e Segal (2004); consultar também Abbott (2008) para contra-argumentos, e o verbete sobre descrições para uma discussão mais extensa, com referências adicionais).

Como a rejeição da distinção de definido/indefinido afetaria o debate se instâncias de descrições definidas referem? Bem, se fosse claro que descrições indefinidas não podem se referir, e se estas fossem equivalentes semânticos das descrições definidas, então teríamos razões adicionais para pensar que descrições definidas não podem ser usadas para referir também. Por outro lado, alguns argumentaram que descrições indefinidas podem referir, ainda que estes argumentos sejam muito controversos (consultar Chastain (1975) e Fodor e Sag (1982); novamente consultar o verbete sobre descrições para discussões e referências adicionais). Mesmo se fôssemos rejeitar tal visão, defensores da teoria unitária ainda tipicamente concederiam que existe uma diferença pragmática, marcada pelo uso de "o" em oposição a "um". Assim, para o defensor mais firme da referência das instâncias de descrições definidas, estaria disponível o argumento de que tais marcadores pragmáticos podem desempenhar uma função significativa na regulação do uso referencial das descrições – apesar destes marcadores não estarem refletidos no significado do que é dito ou asserido. Obviamente, nada disto vai ser resolvido aqui.

5. Modelos da Referência Linguística

Observando como os filósofos tentaram explicar as propriedades referenciais de vários tipos de termos linguísticos, efetivamente apresentamos quatro maneiras distintas de conceber como termos referenciais podem referir a objetos

particulares e individuais no mundo. Em outras palavras, apresentamos quatro modelos para a metasemântica dos termos referenciais:

1. No modelo descritivista, as palavras referem por estarem associadas a um conteúdo descritivo específico, o qual serve para identificar um objeto particular ou individual como o referente.

2. No modelo causal, as palavras referem por estarem associadas à cadeias de uso, as quais levam de volta a um uso inicial ou "batismo" do referente.

3. No modelo do caráter, as palavras referem em virtude de estarem associadas a regras regulares da referência. Regras paradigmáticas deste tipo aludirão a elementos reiteráveis do contexto, identificando quais destes elementos é o referente para cada tipo de termo.

4. No modelo intencionalista, as palavras referem em virtude de serem usadas, intencionalmente, para referir a objetos particulares. Em outros termos, as palavras referem em virtude de serem proferidas como parte de um ato intencional complexo, o qual, de alguma forma, visa objetos particulares ou individuais.

Como já deveria estar claro, estes quatro modelos não são estritamente excludentes uns dos outros. Alguém poderia, por exemplo, pensar no descritivismo clássico – o qual apela para os estados mentais para associar a descrição particular com uma instância referencial do termo – como uma versão particular do intencionalismo. Presumidamente, falantes manifestarão apenas o tipo certo de estado mental quando estão intencionalmente usando um termo referencial para referir. Da mesma forma, alguém poderia conceber a teoria causal como um caso especial da teoria do caráter: nos casos de termos como nomes, a regra invariável de uso associada com um nome particular deve ser explicada em termos de uma cadeia de usos que leva de volta a um objeto ou indivíduo específico. De fato, se alguém está disposto a permitir que os caracteres de certos tipos de termos, como os demonstrativos, aludem às intenções referenciais do falante, então os teóricos do caráter podem aceitar um certo tipo de intencionalismo também. Ao fazer isto, entretanto, não fica claro o que sobra do modelo original do caráter, com a sua dependência de aspectos publicamente identificáveis do contexto como motores principais das regras de referência.

São dignas de nota algumas divisões claras entre estas opiniões. Primeiro, existe uma divisão entre as visões segundo as quais os termos referenciais por si mesmos referem, e assim podemos usá-los para referir porque eles

previamente já referem, e visões segundo as quais somos nós que referimos, e as instâncias linguísticas de referência seguem apenas derivativamente. A teoria causal e certas versões não-clássicas do descritivismo são instâncias paradigmáticas do primeiro tipo de visão, ao passo que o modelo intencionalista é uma instância paradigmática do último – juntamente com descritivismo clássico de Frege e Russell. Devido à sua inerente flexibilidade, o modelo do caráter é mais difícil de localizar neste espectro. As regras de Kaplan para os verdadeiros indexicais parecem atravessar esta divisão: instâncias indexicais adquirem a sua referência porque são proferidas em um contexto particular, presumidamente de maneira intencional. Mas elas adquirem a referência sem ter qualquer relação com o estado mental do falante. As coisas se tornam ainda mais complicadas quando estendemos a teoria do caráter da seguinte forma: os caracteres dos demonstrativos são dependentes da intenção, ao passo que os caracteres dos nomes são constantes e independentes do contexto. Neste caso, a teoria do caráter começa a parecer pluralista; alguns termos referenciais adquirem a sua referência em virtude de nós os usarmos para referir, enquanto usamos outros para referir em virtude deles previamente já referirem.

Uma segunda divisão importante diz respeito à extensão da dependência que a referência pode ter em relação a aspectos potencialmente idiossincráticos e relativamente inacessíveis do estado mental do falante. Como vimos acima, os descritivistas clássicos e os intencionalistas permitem uma forte dependência em relação aos estados mentais dos falantes, e isso pode levar a algumas previsões estranhas acerca da referência. Griceanos, e teóricos da restrição como Reimer, forneceram maneiras de tentar minimizar estes resultados. Porém, alguém também poderia temer que permitir que a referência dependa do estado mental do falante prejudica a nossa habilidade de explicar como nós usamos termos referenciais para transmitir nossos pensamentos direcionados a objetos – em outras palavras, para nos comunicarmos. No modelo causal, por exemplo, literalmente fornecemos modos de pensar sobre o mesmo objeto mediante a transmissão de um nome. Isto torna a explicação da comunicação bastante simples, supondo, é claro, que estejamos consistentemente aptos a reconhecer qual nome foi proferido em uma dada ocasião. O problema de nomes com múltiplos portadores significa que isso pode não ser trivial; mas pelo menos a recompensa parece significativa. Alternativamente, de acordo com a teoria kaplaniana dos verdadeiros indexicais, nós podemos explicar a nossa habilidade de coordenar nosso pensamento sobre lugares particulares, tempos, e indivíduos em virtude do fato de que o lugar, tempo, e lugar do proferimento são são típica-

mente conhecimentos públicos, igualmente disponíveis a cada membro da conversação. A questão é se tal análise está disponível para termos como os indexicais impuros.

Em terceiro lugar, finalmente, existe uma divisão importante entre modelos de referência que são expansivos, em oposição aos restritos. Em suas formas mais amplas, tanto o intencionalismo como o descritivismo são expansivos; eles pretendem ser aplicados igualmente bem a qualquer termo referencial. O modelo causal é um modelo restrito da referência; ele pretendeu apenas ser aplicado à referência dos nomes. Uma vez mais, mostra-se difícil de encontrar o lugar do modelo de caráter; o quanto ele é restrito ou amplo dependerá de quanto estamos dispostos a nos desviar em nossa teorização a partir do caso paradigmático dos verdadeiros indexicais. As teorias expansivas cumprem a promessa de fornecer uma explicação substantiva da relação referencial. Isto é, as teorias expansivas podem alegar que a razão pela qual as instâncias de uso dos termos referenciais desempenham certo tipo de função na explicação da verdade ou falsidade de um proferimento é que certo tipo de relação substantiva obtém entre esta instância de uso e alguns objetos no mundo. Teóricos estritos, ao contrário, terão que permitir diferentes tipos de relações que sirvam para fixar a referência de distintos tipos de termos referenciais, e até mesmo talvez diferentes usos do mesmo termo. Assim, eles terão de dizer que o que unifica os termos referenciais não é que, quando são bem-sucedidos em referir, ocorre um tipo particular de relação entre a instância de uso e algum objeto ou indivíduo, mas sim que eles cumprem algum papel funcional particular na determinação das condições de verdade ou conteúdo assertivo do proferimento. Em outras palavras, por trás do debate sobre qual modelo adotar está uma questão muito maior, e ainda não resolvida: quais são as nossas ambições explicativas, ao propor uma teoria da referência em primeiro lugar?

6. Visões Negativas da Referência

Até aqui, este artigo tratou do que pode-se denominar visões "positivas" da referência. Assume-se que a referência, construída como uma relação entre partes da linguagem e partes da realidade, é uma relação genuína, substantiva, digna de escrutínio filosófico – ou talvez uma família de tais relações. Explicações são então oferecidas a respeito do que constitui esta ligação. Alguns filósofos tentaram desenvolver a partir desta ligação conclusões mais substantivas a respeito da natureza do conhecimento e da realidade (consultar, em particular, Putnam (1981); consultar também o verbete sobre ceticismo e externalismo). No

entanto, nem todos os filósofos estão tão convencidos a respeito da significação teórica da referência, ou mesmo da tese de que existe uma relação, ou uma família de relações substantivas, a serem explicadas. Para terminar, discutiremos brevemente várias visões "negativas" da referência.

W.V.O. Quine é provavelmente o mais famoso cético a respeito da referência. Em Quine (1960), ele argumentou que a referência é inerentemente indeterminada ou "inescrutável". Com isso, Quine quis dizer que não existe matéria de fato a respeito da referência de nossas palavras. Isto não quer dizer que nossas palavras se referem a algo, mas que não somos capazes determinar o que é este algo. Pelo contrário, simplesmente não existe tal coisa que seja "aquilo que é referido por nossas palavras". Mesmo assim, Quine não vai tão longe a ponto de dizer que nossas palavras não referem em sentido algum. A sua visão é de que faz sentido falar da referência apenas de forma relativa a algum propósito que podemos ter ao atribuímos referência a essas palavras. O argumento de Quine para a tese da inescrutabilidade envolve uma aplicação da tese de que teorias empíricas são subdeterminadas pelas suas evidências de apoio. Pois para qualquer corpo de evidência que podemos ter sobre os falantes de uma dada linguagem, sempre haverá um certo número de teorias competidoras a respeito da referência dos falantes. Tais teorias serão empiricamente equivalentes, igualmente consistentes com os dados empíricos. Uma teoria poderia dizer que, na linguagem em questão, "gavagai" se refere a coelhos; uma outra que se refere a partes não destacadas de coelhos; uma terceira poderia dizer que se refere a intervalos de tempo de coelhos. Pode-se aplicar a visão de Quine sobre a subdeterminação aplicada ao próprio idioma da falante. O resultado é que a evidência disponível não torna mais obrigatória para o falante a conclusão de que por "coelho" ele quer dizer coelhos do que a conclusão de que por "coelhos" ela quer dizer partes não destacadas de coelhos ou intervalos de tempo de coelhos. Se um outro falante a observasse utilizando a palavra "coelho", a evidência que ela proporciona dará tratamento igual a todas as três teorias, assim como a muitas outras. Assim, de acordo com Quine, para qualquer corpo de evidência empírica dado, haverá inúmeras teorias competidoras a respeito de qual é a referência das palavras. E não haverá maneira de resolver a disputa entre estas teorias baseando-se em princípios²³.

Um problema relacionado e muito debatido é o problema da multiplicidade, levantado por Peter Unger (1980) e Geach (1980). Basicamente, a ideia é

²³ Consultar o verbete sobre Quine para discussões adicionais.

que para muitos, provavelmente maioria, dos objetos físicos existe um número de candidatos igualmente convincentes para os seus limites, e assim para a sua constituição física. Nos termos com os quais lidamos aqui, há vários candidatos igualmente convincentes para ser o referente de um termo referencial singular, como "isto", quando este termo está sendo usado para alvejar o objeto físico. Sem alguma forma de selecionar algum destes como de alguma forma privilegiados, podemos perguntar se devemos realmente pensar em tais termos como referenciais – a menos, é claro, que estejamos felizes em aceitar que termos referenciais aparentemente singulares de fato referem a conjuntos de objetos, objetos vagos, ou assemelhados²⁴.

As visões instrumentalistas de Davidson sobre a referência são ainda mais radicais. Davidson (1984) alega que a referência é uma noção teoricamente vazia: ela é de absolutamente nenhum uso em uma teoria semântica, ou uma teoria do significado estrito e literal. A sua base para endossar esta posição é sua convicção de que nenhuma explicação substantiva da referência é possível. O problema é que quaisquer explicações deste tipo deveriam ser dadas em termos não-linguísticos, mas nenhuma explicação deste tipo é possível, alega Davidson. Como Davidson coloca: se o nome "Kilimanjaro" se refere a Kilimanjaro, então sem dúvida existe alguma relação entre os falantes do inglês (ou swahili), a palavra, e a montanha. Mas é inconcebível que alguém seja capaz de explicar esta relação sem primeiro explicar a função das palavras nas sentenças; e se isto é assim, não existe possibilidade de explicar a referência diretamente em termos não-linguísticos. Entretanto, isto não significa que não há esperança para a semântica. Muito pelo contrário, na visão de Davidson, uma teoria da verdade para uma linguagem é ao mesmo tempo uma teoria do significado para tal linguagem. O ponto aqui é que uma não tem lugar para a noção da referência per se em uma teoria davidsoniana do significado; ao contrário, ela assimila para dentro de uma teoria da verdade linguística o que podemos estar tentados a denominar de "referência". Isto poderia, entretanto, ser compatível com os tipos de abordagens pluralistas da referência que examinamos acima.

Similares em espírito à visão de Davidson são as teorias deflacionistas da referência – que não devem ser confundidas com deflacionismo sobre verdade, embora sejam frequentemente motivadas pelo desejo de endossar esta última teoria. Deflacionistas sobre verdade tipicamente alegam que asserir que

²⁴ Consultar também Lewis (1993) para uma discussão clássica acerca destes assuntos, e o verbete sobre o problema dos múltiplos para elaboração posterior destas opções.

um enunciado é verdadeiro é o mesmo que asserir o enunciado ele mesmo. Deflacionários sobre referência, por outro lado, alegam que não existe na noção de referência nada além do que é capturado por exemplificações de esquemas como: "a" se refere a a. Tal esquema gera alegações como de que "Frege" se refere a Frege, que podem parecer perfeitamente plausíveis até considerarmos o problema dos nomes com múltiplos portadores ou os indexicais, tanto puros como impuros. Em tais casos, o deflacionista poderia então optar por um esquema um pouco mais complicando, incorporando a noção de uso de um termo. Mas então precisamos perguntar o que é significado por "uso" nestas circunstâncias, e uma vez que começemos a responder este tipo de questão se torna crescentemente nebuloso se uma abordagem verdadeiramente deflacionista pode de fato ser preservada²⁵.

Apesar das inúmeras preocupações que motivam essas visões "negativas" da referência, preocupações que muito poderiam fazer-nos rever a ideia de que existe alguma relação que merecer se denominada de "referência", a natureza da relação entre linguagem e realidade permanece um dos tópicos mais mencionados e vigorosamente debatidos da filosofia da linguagem. O que resta a ser visto é se, e em que medida, este trabalho pode provar ser útil para o entendimento de uma noção mais ampla de referência, ou de uma noção mais ampla da relação entre instâncias representacionais e o mundo à nossa volta, a qual é igualmente acolhida tanto pela estética, ética e filosofia da mente como pela filosofia da linguagem.

Referências

- Abbott, B., 2008. "Issues in the Semantics and Pragmatics of Definite Descriptions in English," in J. Gundel & N. Hedberg (eds.), *Reference: Interdisciplinary Perspectives*, Oxford: Oxford University Press.
- Bach, K., 1992. "Intentions and Demonstrations," *Analysis*, 52(3): 140–146.
- , 2017. "Reference, Intention, and Context: Do Demonstratives Really Refer?," in M. de Ponte and K. Korta (eds.), *Reference and Representation in Thought and Language*, Oxford: Oxford University Press.
- Brandom, R., 1994. *Making it Explicit*, Cambridge MA: Harvard University Press.

²⁵ Para explicações deflacionárias da referência ver, por exemplo, Brandom (1994), Horwich (1998), e Field (2001).

- Burge, T., 1973. "Reference and Proper Names," *Journal of Philosophy*, 70(14): 425–439.
- Chastain, C., 1975. "Reference and Context," in K. Gunderson (ed.), *Language, Mind and Knowledge (Minnesota Studies in the Philosophy of Science: Volume VII)*, Minnesota: University of Minnesota Press.
- Cohen, J. and E. Michaelson, 2013. "Indexicality and the Answering Machine Paradox," *Philosophy Compass*, 8(6): 580–592.
- Davidson, D., 1984. "Reality Without Reference" in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.
- Devitt, M., 1997. "Meaning and Psychology: A Response to Richard," *Noûs*, 31(1): 115–131.
- Dickie, I., 2015. *Fixing Reference*, Oxford: Oxford University Press.
- Donnellan, K., 1966. "Reference and Definite Descriptions," *Philosophical Review*, 75(3): 281–304.
- , 1968. "Putting Humpty Dumpty Together Again," *Philosophical Review*, 77(2): 203–215.
- , 1970. "Proper Names and Identifying Descriptions," *Synthese*, 21(3–4): 335–358.
- Evans, G., 1973. "The Causal Theory of Names," *Proceedings of the Aristotelian Society, Supplementary Volume 47*: 187–208.
- , 1982. *The Varieties of Reference*, Oxford: Oxford University Press.
- Fara, D. Graff, 2015. "Names are Predicates," *Philosophical Review*, 124(1): 59–127.
- Field, H., 2001. *Truth and the Absence of Fact*, Oxford: Oxford University Press.
- von Fintel, K., 2004. "Would you Believe It? The King of France is Back! (Presuppositions and Truth-Value Intuitions)," in A. Bezuidenhout and M. Reimer (eds.), *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions*, Oxford: Oxford University Press.
- Fodor, J.D., and I. Sag, 1982. "Referential and Quantificational Indefinites," *Linguistics and Philosophy*, 5(3): 355–398.
- Frege, G., 1892. "On Sense and Reference," in P. Geach and M. Black (eds.) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell (1952).

- García-Carpintero, M., 1998. "Indexicals as Token-Reflexives," *Mind*, 107(427): 529–564.
- Geach, P., 1969. "The Perils of Pauline," *Review of Metaphysics*, 23(2): 287–300.
- , 1980. *Reference and Generality*, Ithaca: Cornell University Press.
- Gray, A., 2014. "Name-Bearing, Reference, and Circularity," *Philosophical Studies*, 171(2): 207–231.
- , 2015. "Lexical Individuation and Predicativism About Names," *Thought*, 4(2): 113–123.
- Grice, P., 1957. "Meaning," *Philosophical Review*, 66(3): 377–388.
- Horwich, P., 1998. *Meaning*, Oxford: Oxford University Press.
- Jeshion, R., 2015. "Names Not Predicates," in A. Bianchi (ed.), *On Reference*, Oxford: Oxford University Press.
- Kaplan, D., 1978. "Dthat," in P. French, T. Uehling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language*, Minneapolis: University of Minnesota Press.
- , 1989a. "Afterthoughts," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press.
- , 1989b. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press.
- , 1990. "Words," *Proceedings of the Aristotelian Society*, 64: 93–119.
- King, J., 2013. "Supplementives, the Coordination Account, and Conflicting Intentions," *Philosophical Perspectives*, 27(1): 288–311.
- , 1990. "Speaker Intentions in Context," *Noûs*, 48(2): 219–237.
- Kripke, S., 1972. *Naming and Necessity*, Cambridge, MA: Harvard University Press.
- , 1977. "Speaker's Reference and Semantic Reference," *Midwest Studies in Philosophy*, 2: 255–76.
- Lewis, D., 1993. "Many, but Almost One," in J. Bacon (ed.), *Ontology, Causality and Mind: Essays in Honour of D M Armstrong*, Cambridge: Cambridge University Press.
- Ludlow, P. and G. Segal, 2004. "On a Unitary Semantical Analysis for Definite and Indefinite Descriptions," in A. Bezuidenhout and M. Reimer (eds.), *Descriptions*

and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions, Oxford: Oxford University Press.

Mackay, A., 1968. "Mr. Donnellan and Humpty Dumpty on Referring," *Philosophical Review*, 77(2): 197–202.

Marcus, R. Barcan, 1947. "The Identity of Individuals in a Strict Functional Calculus of Second Order," *Journal of Symbolic Logic*, 12(1): 12–15.

—, 1961. "Modalities and Intensional Languages," *Synthese*, 13(4): 303–322.

—, 1993. *Modalities*, Oxford: Oxford University Press.

McGinn, C., 1981. "The Mechanism of Reference," *Synthese*, 49(2): 157–186.

Meinong, A., 1904. "The Theory of Objects," in Meinong (ed.) *Untersuchungen zur Gegenstandstheorie und Psychologie*, Barth: Leipzig.

Mill, J. S., 1867. *A System of Logic*, London: Longmans.

Mount, A., 2008. "Intentions, Gestures, and Salience in Ordinary and Deferred Demonstrative Reference," *Mind and Language*, 23(2): 145–164.

Neale, S., 2004. "This, That, and The Other," in A. Bezuidenhout and M. Reimer (eds.), *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions*, Oxford: Oxford University Press.

Perry, J., 2001. *Reference and Reflexivity*, Stanford: CSLI Publications.

Pelczar, M. and Rainsbury, J., 1998. "The Indexical Character of Names," *Synthese*, 114(2): 293–317.

Predelli, S., 2002. "Intentions, Indexicals, and Communication," *Analysis*, 62(4): 310–316.

Quine, W.V.O., 1960. *Word and Object*, Cambridge MA: MIT Press.

Radulescu, A., 2018. "The Difference Between Indexicals and Demonstratives," *Synthese*, 195(7): 3173–3196.

—, forthcoming. "A Defense of Intentionalism about Demonstratives," *Australasian Journal of Philosophy*, first online 15 January 2019. doi: 10.1080/00048402.2018.1521854

Rami, D., 2014. "The Use-Conditional Conception of Proper Names," *Philosophical Studies*, 168(1): 119–150.

Recanati, F., 1997. *Direct Reference: From Language to Thought*, Oxford: Blackwell.

Reichenbach, H., 1947. *Elements of Symbolic Logic*, New York: Dover.

- Reimer, M., 1991. "Demonstratives, Demonstrations, and Demonstrata," *Philosophical Studies*, 63(2): 187–202.
- , 1992. "Three Views of Demonstrative Reference," *Synthese*, 93(3): 373–402.
- Russell, B., 1905. "On Denoting," *Mind*, 14(56): 479–93.
- , 1911. "Knowledge by Acquaintance and Knowledge by Description," *Proceedings of the Aristotelian Society*, 11: 108–128.
- Schiffer, S., 1981. "Indexicals and the Theory of Reference," *Synthese*, 49: 43–100.
- Schoubye, A., 2011. *On Describing*, Ph.D. Thesis, Department of Philosophy, St. Andrews University.
- Searle, J., 1983. *Intentionality*, Cambridge: Cambridge University Press.
- Soames, S., 2002. *Beyond Rigidity*, Oxford: Oxford University Press.
- Speaks, J., 2016. "The Role of Speaker and Hearer in the Character of Demonstratives," *Mind*, 125(498): 301–229.
- , 2017. "A Puzzle About Demonstratives and Semantic Competence," *Philosophical Studies*, 174(3): 709–734.
- Stebbing, S., 1943. *A Modern Elementary Logic*, London: Methuen.
- Stine, G., 1978. "Meaning Other Than What We Say and Referring," *Philosophical Studies*, 33(4): 319–337.
- Stokke, A., 2010. "Intention-Sensitive Semantics," *Synthese*, 175(3): 383–404.
- Strawson, P., 1950. "On Referring," *Mind*, 59(235): 320–44.
- , 1959. *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen.
- Unger, P., 1980. "The Problem of the Many," *Midwest Studies in Philosophy*, 5(1): 411–468.
- Unnsteinsson, E., forthcoming. "The Edenic Theory of Reference," *Inquiry: An Interdisciplinary Journal of Philosophy*, first online 21 March 2018. doi: 10.1080/0020174X.2018.1446050
- Wittgenstein, L., 1958. *The Blue and the Brown Books*, Oxford: Basil Blackwell.
- Wettstein, H., 1984. "How to Bridge the Gap between Meaning and Reference," *Synthese*, 58(1): 63–84.

(III) A Metafísica da Causação*

Autor: Jonathan Schaffer

Tradução: Bruno Borges & Rodrigo Cid

Revisão: Pedro Merluzzi

Como deve ser um mundo para que ele tenha relações causais? Quando a bola branca encaçapa a bola nove no canto, em virtude de que isso é um caso de causação?

Questões sobre a metafísica da causação podem ser utilmente divididas em questões sobre os relata causais e questões sobre a relação causal. Questões sobre os relata causais incluem as questões de (1.1) se eles estão no espaço-tempo (imanência), (1.2) o quão refinados são (individualização) e (1.3) quantos existem (aridade). Questões sobre a relação causal incluem as questões sobre (2.1) como as sequências causalmente relacionadas e causalmente não relacionadas diferem (conexão), (2.2) como as sequências relacionadas como da causa para o efeito diferem daquelas relacionadas como do efeito para causa ou daquelas como efeitos conjuntos de uma causa comum (direção), e (2.3) como todas as sequências envolvendo causas diferem daquelas que envolvem meras condições de fundo (seleção).

*

Tradução do verbete "The Metaphysics of Causation" de Jonathan Schaffer publicado pela *Stanford Encyclopedia of Philosophy* (Edição do Outono de 2016), edição de Edward N. Zalta, URL = <https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/>. Publicado com a autorização do editor Prof. Dr. Edward Zalta.

The following is the translation of the entry on "The Metaphysics of Causation" by Jonathan Schaffer, in the *Stanford Encyclopedia of Philosophy*. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/causation-metaphysics/>. We'd like to thank the Editors of the *Stanford Encyclopedia of Philosophy*, mainly Prof. Dr. Edward Zalta, for granting permission to translate and publish this entry.

Os filósofos, é claro, têm discordado sobre todas essas questões. A seguir, examinarei alguns dos principais argumentos da literatura.

1. Relata

Questão: O que são relata causais? Quando a bola branca encaçapa a bola nove, quais são os termos dessa relação causal? Uma abordagem dos relata causais deve revelar que tipo de coisa eles são, quantos deles existem e qual o papel de cada um. Em suma, deve revelar sua *categoria, número e função*.

Opções: A visão padrão dos relata causais é que eles são da categoria de *evento* e que seu número é *dois*, nos papéis de *causa* e *efeito*. Assim, na visão padrão, quando a bola branca atinge a bola nove na caçapa do canto, se diz que é um (efetivo) evento e_1 da bola branca atingindo na bola nove, e um (efetivo, distinto) evento e_2 da bola nove caindo na caçapa do canto, desse modo e_1 é causa e e_2 efeito. Em suma, a visão padrão sustenta que os relata causais são um par de eventos.

A visão padrão, claro, tem sido contestada em todos os aspectos. Quanto à categoria, enquanto a visão padrão apresenta os relata causais como eventos (Davidson 1980a e 1980b, Kim 1973, Lewis 1986b), também se encontra apoio considerável para pensá-los como *fatos* (Bennett 1988, Mellor 1995), e suporte ocasional para outras entidades como *características* (Dretske 1977), *tropos* (Keith Campbell 1990), *estados de coisas* (Armstrong 1997), *situações* (Menzies 1989a) e *aspectos* (Paul 2000). As alianças são ainda mais complicadas, por desentendimentos sobre o que são eventos, fatos e essas outras criaturas.

Quanto ao número e ao papel, enquanto a visão padrão numera os relata causais em dois (Davidson 1980b, Mackie 1965, Lewis 1986a), encontra-se algum suporte para visões contrastivas, com *três* e até *quatro* relata (Hitchcock 1996, Woodward 2003, Maslen 2004, Schaffer 2005, Menzies 2007, Northcott 2008, Weslake forthcoming), com o termo(s) adicional desempenhando os papéis de *alternativa causal* e/ou *diferença efectual*. Também se encontra algum suporte para os relata adicionais de um tipo diferente, incluindo *descrições* (Anscombe 1975, McDermott 1995), *modelos* (Menzies 2004, Halpern e Pearl 2005, Hitchcock 2007a) e / ou *estados padrão* (Menzies 2004, McGrath 2005, Hall 2007, Hitchcock 2007a, Halpern 2008). (A seguir, reservarei “o relata” para eventos ou fatos, ou tudo mais, e usarei “relata secundários”

quando desejar falar da perspectiva das descrições, dos modelos ou dos padrões como relata adicionais. Isto é puramente por conveniência expositiva.)

Categoria: O que é a categoria de relata causal? Que tipo de coisa são eles? Uma abordagem da categoria dos relata deve, primeiro, determinar quais características diferenciam os eventos dos fatos e dos outros, e, em seguida, identificar quais características os relata devem ter.

Na prática, existem duas principais características diferenciadoras que são invocadas. A primeira é a *imanência*. Os eventos e os outros são geralmente considerados imanentes, enquanto os fatos são geralmente considerados transcendententes. Isto é, o evento de Brutus esfaqueando César é algo concreto que ocorre em uma localização particular no espaço-tempo (o Senado, nos idos de março), enquanto o fato de Brutus ter esfaqueado César é algo abstrato e não situado no espaço-tempo. Surge então a questão de saber se a causação requer relata imanentes ou transcendententes.

O segundo principal diferenciador que é invocado é a *individuação*. Por vezes (embora nem sempre) os eventos são considerados relativamente grosseiros, enquanto que os fatos e os outros são considerados relativamente refinados. Por exemplo, o evento de John dizer "olá" pode ser considerado o mesmo que o de John dizendo "olá" em voz alta, enquanto o fato de John dizer "olá" é diferente do fato de que John diz "olá" em voz alta. Surge então a questão de saber se a causação requer relata grosseiros ou refinados.

Felizmente, questões sobre a verdadeira metafísica de eventos, de fatos e de outros candidatos podem ser adiadas aqui, e as questões sobre a imanência e a individuação dos relata causais podem ser tratadas diretamente. Assim, a questão da categoria dos relata pode, na prática, ser utilmente substituída por duas questões: se os relata são imanentes (Seção 1.1), e o quão refinado eles são individuados (Seção 1.2).

Número e papel: Qual é o número e o papel dos relata causais? Quantos relata causais existem, e que tipo de trabalho eles fazem? Uma abordagem do número e do papel dos relata deve primeiro formular determinantes gerais da aridade das relações, e então aplicar esses determinantes à causação.

A visão de que existem dois relata é amplamente assumida, mas raramente defendida. Três principais alternativas foram exploradas envolvendo contraste. A primeira dessas alternativas, inspirada no trabalho de Van Fraassen (1980) sobre explicação contrastiva, toma as relações causais como incluindo

uma *diferença efectual*. Nesta visão causal, as relações têm a forma: c causa e ao invés de e^* . A segunda alternativa principal, baseada em Hitchcock (1993, 1995a, 1996), trabalha com causação probabilística, toma as relações causais como incluindo uma *alternativa causal*. Nisto, relações causais têm a forma: c ao invés de c^* causa e . A terceira alternativa principal, defendida por Schaffer (2005), inclui tanto uma alternativa causal quanto uma diferença efectual e , assim, numera o relata em quatro, produzindo a forma: c ao invés de c^* causa e ao invés de e^* . Surge então a questão de saber se contrastes (por causa e / ou efeito) ajudam a resolver problemas e paradoxos (Seção 1.3).

Três alternativas principais foram exploradas, postulando relata secundários. A primeira delas, inspirada na afirmação de Anscombe (1975) de que a causação é uma relação intensional, toma a causação como sendo relativa às descrições dos relata primários. Nessa visão, as relações causais têm a forma: c causa e relativo à D , onde D é um par ordenado de descrições (para c e para e). A segunda dessas alternativas, resultante especialmente do trabalho de Pearl (2000) sobre modelagem causal, trata a causação como relativa a certo tipo de representação matemática. Sob esse ponto de vista, as relações causais têm a forma: c causa e relativo a M , onde M é um modelo causal adequado da situação. A terceira dessas alternativas, que tem raízes no tratamento da causação com leis de Hart e Honore (1985), trata a causação como relativa a estados-padrão, que codificam os estados que são considerados “normais” e “naturais”, em oposição aos estados desviantes. A versão mais simples dessa visão tem a forma: c causa e relativo a N , onde N é um par ordenado de resultados naturais (referentes a c e e). Essas visões são todas compatíveis. Pode-se, por exemplo, sustentar que as relações causais têm a forma: c causa e relativo a D , M e N . (De fato, uma área de pesquisa ativa diz respeito à combinação da modelagem causal com os padrões: veja Blanchard e Schaffer (no prelo) para uma crítica. Em seguida, surge a questão de saber se é necessário algum relata secundário, ou se todos eles constituem uma perda objetável de objetividade, ou um afastamento objetável da suposta “óbvia binariedade” da causação (Seção 1.3).

Pressuposições: Tanto a disputa sobre a categoria quanto sobre os números e o papel dos relata causais envolvem pressupostos de *singularidade*. Quanto à categoria, a disputa pressupõe que exista uma categoria única de entidade a partir da qual todos os relata causais devem ser traçados. Contudo, pode-se argumentar que a linguagem comum permite que a relação seja descrita em formas de eventos (imperfeita nominal), de fatos (perfeita nominal) e outras

(Mackie 1974, Vendler 1984, Bennett 1988). Por que não tomar a linguagem comum em sua expressão ordinária, e deixar mil relata florescerem?

Quanto ao número e ao papel, a disputa pressupõe que há um único número que é a aridade. No entanto, pode-se argumentar, novamente, que a linguagem comum permite atribuições causais com e sem alternativas causais ou diferenças efetivas (Hitchcock, 1996). Por que não tomar a linguagem comum em sua expressão, e tomar a causação como multigrau?

Existem dois argumentos principais em defesa da singularidade, o primeiro é o de evitar a ambiguidade (Menzies, 1989a). Se houvesse quatro opções para dois relata, poderia parecer que haveria $2^4=16$ relações "causais" (ou mais se houvesse mais opções e / ou relata e / ou aridade). Dito isso, não está claro por que não poderia haver uma única relação causal (univocamente denotada por "causação") que permitisse diferentes tipos de relação. A relação de identidade, por exemplo, pode relacionar itens de qualquer categoria ontológica.

O segundo argumento para a singularidade é que ela impede uma *harmonia* misteriosa (Mellor, 1995). Se houvesse uma pluralidade de causas de eventos e causas de fatos e afins, alguma harmonia metafísica seria necessária entre eles, pois certamente eles não poderiam entrar em conflito. Isto é, parece que o evento de uma bola branca atingindo a bola nove e o fato de a bola branca ter atingido a bola nove devem ter efeitos comparáveis. Mas sem uma única relação causal subjacente única, parece não haver nada que mantenha esses efeitos alinhados. Dito isso, talvez uma pluralidade de relações causais pudesse ser harmonizadas, desde que (i) uma fosse fundamental e as outras derivadas, ou (ii) todas fossem derivadas de uma base não causal comum, como as regularidades entre os eventos.

1.1. Imanência

Questão: Os relata causais são imanentes ou transcendententes? Ou seja, são eles concretos e localizados no espaço-tempo, ou abstratos e não-espaço-temporais?

Esta questão está ligada à questão da categoria. Se os relatas são transcendententes, então são fatos. Se eles são imanentes, então são eventos ou um dos outros candidatos, como recursos, tropos ou situações.

Na prática, encontramos dois argumentos principais sobre a questão da imanência. Primeiro, há o argumento de empurrar, que sustenta que os relata

devem ser imanentes para empurrarem as coisas. Em segundo lugar, há o argumento das ausências, que sustenta que os fatos devem ser transcendentais para que as ausências possam figurar em relações causais.

Empurrar: O principal argumento para a imanência é que apenas entidades imanentes podem *interagir*. Esse argumento é bem resumido por um de seus oponentes, Bennett: “Algumas pessoas objetaram que os fatos não são o tipo de item que podem causar coisa alguma. Um fato é uma proposição verdadeira (dizem eles); não é algo *no* mundo, mas sim algo *sobre* o mundo, o que torna categoricamente errado o papel de puxar, enfiar, torcer e dobrar” (1988, p. 22; ver também Hausman 1998). De acordo com o argumento de empurrar, apenas entidades espaço-temporais concretas podem ser causas e efeitos.

Há duas respostas principais ao argumento de empurrar, o primeiro é encontrar entidades imanentes substitutas. Esses substitutos imanentes servem como empurradores e relacionam-se com os fatos causais, embora ainda sejam distintos deles. Bennett, na continuação imediata da citação acima, recruta *objetos* para tal propósito: “Isso se baseia na suposição errônea de que as declarações causais devem relatar relações entre coisas que empurram e forçam. Admito que os fatos não podem se comportar como uma cotovelada nas costas, mas sabemos quais itens desempenham esse papel - a saber, os cotovelos. Em nosso mundo, o empurrar, o enfiar e o forçar são feitos pelas coisas - partículas elementares e agregados delas - e não por qualquer relata da relação causal.” (1988, p. 22) Mellor (1995) oferece uma resposta semelhante, sugerindo *facta* (os veridadores imanentes para os fatos) como a base imanente da causação de fatos.

A segunda resposta ao argumento de empurrar é acusar que ele repousa sobre uma concepção ingênua (pré-humana) de causação exigindo algum tipo de impulso metafísico ou “oomph”. Se a relação causal é uma mera questão de regularidade, por que as regularidades não podem se manter entre os fatos?

Ausência: O principal argumento para a transcendência é que as *ausências* podem estar envolvidas em relações causais. As ausências são consideradas entidades transcendentais. Eles *nada* são, não ocorrem e, portanto, não estão no mundo. Assim, Mellor diz: “Pois o “C” e o “E” em um verdadeiro “E porque C” causal não precisa afirmar a existência de particulares. Eles podem negar isso... Eles são afirmações existenciais negativas, feitas verdadeiras pela não-existência de tais particulares,...” (1995, p. 132) Aqui Mellor está

argumentando que, no caso em que o escalador Don não morre, porque ele não cai, o não caindo e não morrendo de Don estão causalmente relacionados, sem que haja qualquer evento ou outras entidades imanentes para relacionarmos.

Há duas respostas principais ao argumento da ausência, a primeira é negar que as ausências possam ser causais. Nesse sentido, Armstrong afirma: “Omissões e semelhantes não fazem parte da verdadeira força motriz da natureza. Toda situação causal se desenvolve como resultado da presença de fatores positivos sozinhos.” (1999, p. 177; ver também Beebe 2004a e Moore 2009) O teórico que nega a causação da ausência pode acrescentar algum codicilo conciliatório no sentido de que as ausências se mantêm em relações com formato causal. Assim, Dowe (2000, 2001) desenvolve uma abordagem de substituto da causação (causação*) para explicar nossas intuições de que as ausências podem ser genuinamente causais.

A segunda resposta ao argumento da ausência é negar que as ausências são transcendentais. Uma maneira de fazer isso seria aceitar a existência de propriedades negativas e pensar em ausências como eventos nos quais um objeto instancia uma propriedade negativa. Assim, a instanciação de Don, a não-queda em t_0 , pode ser contada como um evento imanente, e uma causa do evento ainda mais imanente de sua instanciação do não-morrer em t_1 . Uma segunda maneira de negar que as ausências são transcendentais seria considerar as declarações de ausência meramente como uma maneira de descrever ocorrências, como Hart e Honore recomendam: “O correto aqui é perceber que declarações negativas como “ele não puxou o sinal” são formas de descrever o mundo, assim como asserções afirmativas, mas descrevem-no em *contraste*, e não por *comparação*, como afirmações afirmativas fazem.” (1985, p. 38) Assim, Don não caindo em t_0 pode ser identificado como se agarrando à rocha em t_0 e Don não morrendo em t_1 pode ser identificado com sua sobrevivência em t_1 , cujos eventos são de fato relacionados de modo causal.

1.2. Individuação

Questão: Como os relata causais são individuados? Isto é, se r_1 e r_2 são relata causais, quais são as condições que determinam se $r_1 = r_2$?

Esta questão está relacionada com a questão da categoria. Se os relata são grosseiros, então são eventos de certo tipo. Se são refinados, então são fatos, ou um dos outros candidatos, como recursos, tropos, situações ou então eventos de outro tipo.

Combinando as considerações sobre a imanência e a individuação, pode-se, como uma primeira aproximação, distinguir um quadrado de possíveis visões subjacentes à disputa sobre categoria:

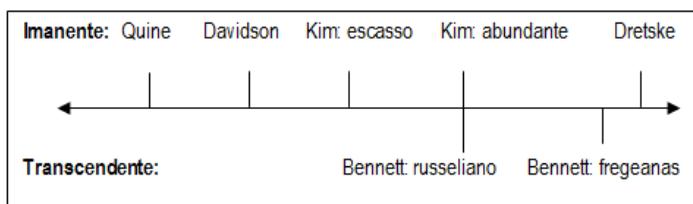
	Grosseiros	Refinados
Imanente	Davidson	Kim, Lewis, Dretske, etc.
Transcendente	[Vazio]	Bennett, Mellor

No quadrante superior direito, encontram-se, ao lado de teóricos de eventos refinados como Kim e Lewis, virtualmente todos aqueles que optaram por uma terceira via, incluindo Dretske, Campbell, Armstrong, Menzies e Paul. De fato, quase todos os teóricos que rejeitaram tanto eventos quanto fatos o fizeram porque consideram que os relata devem ser imanentes e, portanto, não são fatos, porém também devem ser refinados e, portanto, não são eventos.

Mas o quadrado é apenas uma primeira aproximação muito grosseira, porque a granularidade realmente vem em graus. Assim, em uma segunda aproximação, poder-se-ia pensar a granularidade em um continuum, a partir da grosseira visão de Davidson (1980c) na qual os relata causais são individuados por suas causas e efeitos, ao moderado refinamento da visão de Kim (1976) segundo a qual os relata causais são individuados por suas triplas <objetos, propriedade, tempo> associadas, até ao extremo refinamento de Bennett (1988), no qual os relata causais são individuados tão refinadamente quanto proposições.

O continuum de Davidson-Kim-Bennett ainda é apenas uma aproximação, porque realmente existem outros pontos ocupados ao longo de um continuum mais amplo. Quão refinado são, de fato, os eventos de Kim, isso depende de o quão finamente as propriedades são individuadas. Se as propriedades são tomadas em um sentido abundante (individuadas tão finamente quanto predicados, ou pelo menos tão finamente quanto predicados necessariamente coextensivos) então os eventos kimianos são relativamente refinados, ao passo que, se as propriedades são tomadas em um sentido esparso (individuadas pelas juntas da natureza”) então o grão é mais grosso. O quão refinado os fatos de Bennett são, depende de quão finamente as proposições são individuadas. Se proposições são tomadas como Frege as concebe, então os fatos de Bennett são excepcionalmente refinados; enquanto que, se proposições são tomadas como Russell as concebe, então os fatos de Bennett são comparáveis em granularidade aos eventos de Kim com propriedades abundantes. Além disso, a visão de Davidson é superada em grossura pela visão de Quine (1985; aceita por Davidson, 1985) de que os relata causais são individualizados pela região do espaço-tempo. Embora até a variante fregeana da visão de Bennett seja superada em refinamento pela visão de Dretske (1977) na qual até as diferenças focais (como entre Mary beijando John e Mary beijando John) acarretam diferenças no relata. Colocar isso em conjunto produz a seguinte imagem:

Individuação: Grosseiros → Refinados



Outras discriminações ao longo da dimensão de granularidade são, evidentemente, possíveis.

Na prática, encontramos três argumentos principais sobre a questão da individuação. Em primeiro lugar, há o argumento das diferenças causais, que sustenta que o relata deve ser refinado, de modo a marcar diferenças na relação causal. Em segundo lugar, há o argumento da transitividade, que sustenta que o

relata deve ser refinado, para preservar a transitividade. Terceiro, há o argumento da metodologia, que sustenta que o relata deve ser grosseiro por razões gerais de elegância teórica.

Diferenças causais: O primeiro argumento para o refinamento é que diferenças refinadas podem marcar *diferenças causais*. Para dar um exemplo discutido por Davidson, faz sentido dizer: "O colapso foi causado, não pelo fato de que o parafuso cedeu, mas pelo fato de ele ter cedido de forma tão repentina e inesperada" (1980b, p. 161; veja também Kim 1976). Isso sugere que o ferrolho se partindo e o ferrolho se partindo de repente devem diferir como relata causal. Ou para citar um exemplo de Lewis (1986b), John dizendo "olá" deve diferir de John dizendo "olá" em voz alta, já que apenas o primeiro faz com que Fred cumprimente John de volta, e apenas o segundo é causado pelo estado de tensão de John. De acordo com o argumento das diferenças causais, o relata deve ser refinado sob pena de misturar relata causais conflitantes.

Existem três principais respostas ao argumento das diferenças causais, a primeira é que "causa" é ambígua entre causação e explicação. Diz-se que o argumento das diferenças causais apenas introduz diferenças explicativas (Davidson 1980b e 1980d, Strawson 1985). Davidson integra essa resposta a uma abordagem geral de causação e da explicação, na qual a causação é uma relação extensional que se mantém entre os eventos grosseiros, enquanto a explicação é uma relação intencional que se mantém entre os eventos grosseiros sob uma descrição. Assim, o estado de tensão de John faz com que John esteja dizendo "olá", mas a relação explicativa só se aplica quando o "olá" é descrito em termos de seu volume. Nosso juízo de que o estado de tensão de John não "causa" John dizendo o "olá" deve ser rejeitado, por estar ligado ao idioma das explicações.

A segunda resposta ao argumento das diferenças causais é que "... causa..." é um contexto intensional (Anscombe 1975, Achinstein 1975 e 1983, McDermott 1995). Contextos intencionais não licenciam a substituição de termos co-referenciais a salva veritate. Assim, John dizendo "olá" pode se referir ao mesmo evento que John está dizendo "olá" em voz alta, mas substituir a única descrição pelo outro ainda pode mudar o valor de verdade da afirmação causal. Se isso é o caso, diferentes descrições do mesmo relata podem induzir diferenças causais.

A terceira resposta ao argumento das diferenças causais é que ele extrapola. Parece exigir um refinamento além do que muitos de seus proponentes previram. Para dar um exemplo de Achinstein (1983), pode ser

verdade que “a cicuta bebida por Sócrates ao anoitecer causou sua morte”, mas é falso que “a cicuta bebida por Sócrates ao crepúsculo causou sua morte”. Se assim for, o argumento das diferenças causais implica que a cicuta bebida por Sócrates ao anoitecer e a cicuta bebida por Sócrates ao crepúsculo devem diferir como relata causais. Se esse for o caso, então o argumento das diferenças causais requer, em última análise, um extremo refinamento da visão dreetskiana, que alguns podem considerar como em *reductio*.

Transitividade: O segundo principal argumento para o refinamento é que ele preserva a *transitividade*. Para adaptar um exemplo de Woodward (1984; ver também Ehring 1997, Paul 2000), suponha que Tom coloque sal de potássio na lareira (*c*), e então Dick lança um fósforo, o que resulta em um fogo púrpura brilhante na lareira (*d*), que então se espalha e incendeia Harry (*e*). O teórico não refinado parece estar comprometido com a seguinte quebra de transitividade: Tom colocando sal de potássio na lareira faz com que haja um fogo púrpura brilhante: *c* causa *d*; há um fogo roxo aceso na lareira que causa a imolação de Harry: *d* causa *e*; mas Tom colocando sal de potássio na lareira não causa a imolação de Harry: *c* não causa *e*. O teórico refinado pode distinguir *d*₁: o fogo tornando-se roxo na região *r*, de *d*₂: o fogo ardendo em *r*. Agora, *c* causa *d*₁ (não *d*₂), *d*₂ causa *e* (não *d*₁), e assim a inferência transitiva de *c* causando *e* seria bloqueada.

Existem duas respostas principais para o argumento da transitividade, a primeira é de fato aceitar a consequência. Isto é, pode-se aceitar que *c* cause *e*: Tom colocando sais de potássio na lareira realmente causa a imolação de Harry. Nossas intuições contrárias podem ser rejeitadas, como apresentadas acima, ao misturarmos os idiomas causais e explicativos. Ou nossas intuições em contrário podem ser rejeitadas em bases pragmáticas. Como observa Lewis, em um contexto relacionado, estamos propensos a “misturar questões sobre o que é geralmente apropriado com questões sobre o que causou o que”, embora: “todo historiador sabe que as ações geralmente têm consequências não intencionadas e indesejadas”. (2000, pp. 194-5)

A segunda resposta ao argumento da transitividade é negar que a causalidade é transitiva. Parece haver falhas de transitividade que não podem ser resolvidas com refinamento. Para pegar um exemplo de Hall (2000, veja também Hitchcock 2001), suponha que um pedregulho comece a descer a colina em direção à cabeça do caminhante (*c*), o que faz com que o caminhante pule (*d*), e isso faz com que o caminhante sobreviva (*e*). Parece que *c* causa *d* e que *d* causa *e*, no entanto, não parece que *c* causa *e* ou que dividir *d* em diferentes

características, ou aspectos, ou algo semelhante vai ajudar. Se assim for, a transitividade é perdida de qualquer maneira.

Metodologia: O principal argumento para a visão não refinada é que esta é metodologicamente preferível. Quine (1985, p. 167) atribui às concepções refinadas dos relata a postulação de entidades não familiares e pobremente individualizadas, e recomenda a individuação extremamente não refinada (espaço-temporal) como justificar em princípios e familiar aos compromissos ontológicos. E Davidson (1985) considera a visão de Quine como "mais pura" e "melhor" do que a sua própria visão anterior (1980b).

Há duas respostas principais para o argumento da metodologia, a primeira é que algumas das concepções refinadas são perfeitamente justificadas em princípios. Para Kim (1976) concepção refinada de eventos como instanciada de propriedade, por exemplo, oferece um critério preciso para a individuação, a saber:

$(\forall x)(\forall y)$ (Se x e y são eventos, então $x = y$ sse [objeto(x) = objeto(y) & propriedade(x) = propriedade(y) & tempo(x) = tempo(y)]).

(A generalização para as relações n -árias é direta.) E as entidades invocadas (objetos, propriedades e tempos) devem ser perfeitamente familiares para todos, menos para o nominalista mais severo. De fato, o próprio Quine admite não apenas que a concepção refinada de Kim é perfeitamente baseada em princípios, mas até mesmo que ela "ainda poderia ser acomodada na ontologia que tenho aceitado" (1985, p. 167).

A segunda resposta ao argumento da metodologia é que não está claro que essas indecências na individuação ou que a multiplicação de entidades deva contar muito. Quanto à individuação, aceitamos objetos físicos sem princípios claros de individuação, então por que avaliar eventos (ou qualquer que seja o relata) a partir de um padrão mais rigoroso? Quanto à multiplicação, se alguém tem uma visão refinada e redutora (como a visão de Lewis 1986b de eventos como classes transmundiais de regiões), então não há nenhuma multiplicação em sua ontologia básica, já que todos os componentes já existem. E se alguém tiver uma visão refinada e não-redutora, a multiplicação resultante ainda pode ser irrepreensível. O verdadeiro pecado metodológico é multiplicar entidades sem necessidade, portanto, se houver necessidade de uma individuação refinada (como os argumentos das diferenças causais e/ou da transitividade podem fornecer), então a postulação de tais entidades é metodologicamente pura.

1.3. Aridade

Questão: Qual é o número e o papel dos relata causais? Isto é, quantos há, e que tipo de trabalho cada um executa? Esta questão pode ser utilmente dividida em sub-questões, como a de saber se devemos postular contrastes (seção 1.3.1), e se devemos postular qualquer relata secundários, como descrições, modelos ou padrões (seção 1.3.2).

1.3.1. Contrastividade

Quanto aos contrastes, encontramos quatro principais argumentos. Primeiro, há o argumento da forma superficial, que sustenta que os relata causais devem ser dois, para se adequar à forma superficial de enunciados tais como "o curto-circuito causou o fogo". Segundo, há o argumento da determinação, que sustenta que o os relata causais devem ser quatro, para que as relações causais sejam bem definidas. Terceiro e quarto, há os argumentos da imanência revisitada e da individuação revisitada, que sustentam que os relata causais devem ser quatro, para resolvermos os problemas sobre ausências e diferenças causais, revisados acima (Seções 1.1 e 1.2, respectivamente).

Todos os argumentos a serem considerados aqui exigem dois ou quatro relata (não três). De fato, pode-se argumentar contra visões contrastivas de três relações, que elas impedem as *cadeias causais*. Em uma cadeia causal, o efeito no primeiro elo serve como causa no segundo. Para que isso seja possível, causa e efeito devem ser formalmente permutáveis: a mesma estrutura deve flanquear os dois lados da relação. Suponha que o primeiro dominó derrube o segundo, que então derruba o terceiro. O teórico binário pode dizer que c : a queda do primeiro dominó causa d : a derrubada do segundo; e que d , por sua vez, causa e : a derrubada do terceiro dominó. O teórico quaternário pode dizer que c em vez de c^* : o primeiro dominó que permanece em pé causa d em vez de d^* : o segundo dominó permanece em pé; e que d , em vez de d^* , causa e e não e^* : o terceiro dominó permanece em pé. Mas se houvesse contrastes em apenas um lado da relação, então essas cadeias não poderiam ser construídas. As ligações não coincidiriam.

Forma Superficial: O principal argumento para a binariedade é que ela é revelada pela *forma superficial* das asserções causais. As reivindicações causais como "o curto-circuito causou o incêndio" não fazem referência explícita

a quaisquer contrastes. Tais alegações podem ser proferidas do nada (na posição inicial do discurso) e, portanto, também não requerem nenhum ajuste prévio de contraste ou fixação de pressupostos. Esta é, presumivelmente, a raiz da ideia de que a causação é “obviamente binária”. De fato, esse tipo de consideração é mais proeminente em Davidson (1980b), que busca a forma lógica de enunciados de superfície binária. Da mesma forma, Davidson rejeita essa noção próxima de *relevância causal* porque “Não há espaço para um conceito de “causa” que faria da causalidade uma relação entre três ou quatro entidades e não entre duas.” (1993, p. 6)

Existem três respostas principais ao argumento da forma superficial. A primeira é que formas superficiais contrastivas também existem. Por exemplo, também encontramos alegações de que “Pam jogando uma pedra em vez de uma pedrinha causou o estilhaço da janela”, de que “Pam jogando a pedra para a janela se estilhaçar em vez de trincar”, e até a combinação “Pam estaria jogando a pedra ao invés da pedrinha, causou a janela se estilhaçando em vez de trincar.” Assim, a forma superficial pode parecer equivocada (Hitchcock, 1996).

A segunda resposta ao argumento da forma superficial é que esta pode ser uma expressão reduzida de uma forma lógica mais complexa. Por exemplo, “Ann prefere chocolate” pode ser usado como uma expressão reduzida da proposição de que Ann prefere chocolate ao *invés de baunilha*. Aqui o contraste não precisa ser explicitamente articulado, ou mesmo explicitado anteriormente na conversa. Pode-se realizar asserções de preferência como superfície binária do nada. Então, assim como as reivindicações de preferência podem parecer ter uma forma lógica contrastiva sob sua superfície binária, as alegações causais também podem ser consideradas sob essa luz. De qualquer forma, como a história da semântica demonstrou, seria tola mente ingênuo pensar que a linguagem comum exibe sua forma lógica em sua superfície. Nesse sentido, Schaffer (2012) diagnostica uma série de dependências contextuais no discurso causal como sendo devidas à contrastividade implícita.

A terceira resposta principal está no *revisionismo*. Mesmo que provemos que a forma lógica das atribuições causais seja binária, a forma lógica não deve ter a última palavra na metafísica, pois ela pode predicar uma propriedade que temos uma razão teórica para rejeitar. Por exemplo, “A rocha está se movendo” pode parecer predicar a propriedade do movimento absoluto, mas os físicos descobriram que não existe tal coisa. Assim como asserções de movimento apenas fazem sentido metafísico quando relativizadas a um referencial inercial,

pode-se pensar que as afirmações causais só fazem sentido metafísico quando relativizadas a contrastes.

Determinação: O primeiro argumento principal para a quaternicidade é que as relações causais binárias são *mal definidas*. Suponha que Jane fume moderadamente e desenvolva câncer de pulmão. O fumo moderado de Jane causa câncer de pulmão? Hitchcock diz que não há uma resposta determinada a menos que se fixe a alternativa causal: “A solução para este enigma é negar que exista algo como a relevância causal do tabagismo moderado para o câncer de pulmão... Em relação ao tabagismo pesado, é uma causa negativa de (previne) câncer de pulmão; Em relação à abstinência, o tabagismo moderado é uma causa positiva de (causa) câncer de pulmão... Relações de relevância causal positiva ou negativa apenas se mantêm em relação a alternativas específicas.” (1996, p. 402) Um caso paralelo poderia ser construído para a necessidade de dar conta da diferença efectual. Suponha que Pablo esteja escolhendo entre tinta azul, vermelha e verde para sua tela. Será que Pablo escolhendo a tinta azul em vez da vermelha, causa a tela fica azul? Aqui se pode dizer que ainda não há uma resposta determinada. Pablo, ao escolher a tinta azul em vez da vermelha, faz com que a tela seja azul em vez de vermelha, mas não faz com que a tela seja azul em vez de verde. Assim, pode-se concluir que os contrastes são necessários tanto para causa quanto para efeito, a fim de que as afirmações causais tenham valores de verdade determinados.

A principal resposta ao argumento da determinação é que as relações causais binárias são bem definidas, afinal. Essa resposta deve tomar a forma de aplicar uma abordagem binária da causação a casos problemáticos, como os casos de fumar e de pintar, e simplesmente avaliar ser valor de verdade, seja ele qual for. Por exemplo, alguém poderia pensar que uma abordagem direta contrafactual da causação, em que checamos se e ainda teria ocorrido, caso não tivesse ocorrido, simplesmente acarreta que o fumar de Jane causa câncer de pulmão, e que Pablo ao escolher tinta azul, faz a tela ficar azul, ponto final.

Imanência Revisitada: O segundo principal argumento para a abordagem quaternária consiste em *revisitar a imanência* e sustentar que argumentos adicionais tentam reconciliar a imanência com a causação por ausência (Schaffer, 2005). A reconciliação é tentada através de reivindicações de ausência como configuração para contraste das ocorrências associada. Por exemplo, “o jardineiro não molhar as flores fez com que as flores murchassem” deve ser interpretado como: o que o jardineiro fez na realidade (isto é, o evento efetivo de assistir televisão) em vez de regar as flores (o evento não-efetivo que

é a ocorrência associada) fez com que as flores murchassem em vez de desabrocharem. E essa afirmação pode muito bem ser verdadeira. Dessa forma, todos os quatro relatos podem ser tratados como entidades imanentes, e as alegações causais de ausência podem ainda assim ser verdadeiras. Desta forma, a ausência de causação não requer nenhuma provisão especial (o que, como Dowe (2000) explica, não é o caso para qualquer teoria binária).

A principal resposta ao argumento de revisar a imanência é que a imanência não precisa ser revisitada. Nenhuma reconciliação de ausências e imanência é necessária e, portanto, nenhum argumento adicional é necessário. Essa resposta pode tomar a forma de negar a imanência (isto é, abraçar fatos), negar que as ausências são causais ou sustentar que existem ausências imanentes (Seção 1.1).

Individuação revisitada: O terceiro principal argumento para a abordagem quartenária *revisita a individuação* e sustenta que os lugares para argumentos adicionais domam o argumento das diferenças causais (Schaffer, 2005). A preocupação é que o argumento das diferenças causais extrapola, ao exigir que Sócrates *beber cicuta* ao anoitecer e Sócrates beber cicuta *ao anoitecer* sejam diferentes como relatos causais. A domesticação é tentada através do tratamento das diferenças focais como diferenças contrastantes. Assim, “Sócrates *beber cicuta* ao anoitecer” deve ser interpretada como *c*: Sócrates beber cicuta ao anoitecer, em vez de *c**: Sócrates beber vinho ao anoitecer (ou alguma alternativa contextualmente saliente à ingestão de cicuta); enquanto “Sócrates beber cicuta *ao anoitecer*” deve ser interpretado como *c*: Sócrates beber cicuta ao anoitecer, em vez de *c**: Sócrates beber cicuta ao alvorecer (ou alguma alternativa contextualmente saliente a ocorrer ao entardecer). E esses diferentes contrastes podem induzir a diferentes efeitos. Desta forma, diferenças focais podem produzir diferenças causais, sem ter quaisquer implicações para a individuação, muito menos para o extremo refinamento da visão dretskiana.

A principal resposta ao argumento da individuação revisitada é que a individuação não precisa ser revisitada. Não é necessário domar o argumento das diferenças causais e, portanto, nenhum lugar para o argumento adicional é necessário. Essa resposta pode tomar a forma de manter a distinção davidsoniana entre causação e explicação causal, de permitir que a causação seja uma relação intencional, ou de simplesmente aceitar a visão dretskiana da relação (Seção 1.2).

1.3.2. Relata Secundário

Quanto ao relata secundário, encontramos três argumentos principais. Primeiro, há o argumento da objetividade, o qual sustenta que quaisquer relativizações de relações causais como descrições, modelos ou padrões são incompatíveis com a objetividade da causação. Em segundo lugar, há o argumento do sucesso das técnicas de modelagem causal, que sustenta que a causação deve ser relativa a um modelo, de modo que as técnicas de modelagem possam ser exploradas. Terceiro, argumentou-se que a relatividade padrão fornece a melhor explicação para as nossas intuições em vários casos problemáticos. (O argumento da forma superficial (Seção 1.3.1) também pode servir como um quarto argumento contra qualquer relata secundário.)

Objetividade: Deixando de lado questões linguísticas da forma superficial, o principal argumento contra descrições, modelos e / ou padrões como relata secundário é que incluí-los comprometeria a objetividade da causação. Quando a bola branca encaçapa a bola nove, pode parecer que existe uma causa objetiva no mundo. Como escolhemos descrever os eventos, modelar a situação ou rotular determinados resultados como "padrão" ou "desviante" parece ir além do ponto. Por exemplo, se um modelo M é usado no qual sua interação aparece como não-causal, pode-se inferir, *não* que a interação não seja causal em relação a M , mas sim que o próprio M é um modelo ruim dos fatos causais dependentes de modelo.

Existem duas respostas principais (relacionadas) ao argumento da objetividade, a primeira das quais é aceitar a consequência e negar que a causação é objetiva. Tal resposta pode ser parcialmente encorajada pelo pensamento de que a causação parece abandonar a física fundamental (2.1.2). A segunda resposta é sustentar que ainda existe objetividade em uma ampla gama de noções associadas. Assim, Hitchcock invoca a noção de *estrutura causal particular*, para aquilo que os modelos causais representam e sugere que “podemos permitir que os julgamentos da causação particular sejam infectados por critérios pragmáticos sem abandonar a objetividade da causação em geral: A objetividade pode ser retida no nível de estrutura causal simbólica” (2007a, p. 504).

Modelagem Causal: Talvez o argumento principal para a relatividade do modelo comece com o sucesso das técnicas de modelagem causal, como desenvolvido em Spirtes, Glymour e Scheines (1993), Pearl (2000) e Halpern e Pearl (2005), *inter alia*. Essas técnicas forneceram algoritmos de trabalho para a descoberta causal e renderam elegantes abordagens para as relações causais

tipo-nível. É apenas prudente, começa o argumento, buscar uma explicação da causação (real, simbólica) usando tais técnicas. De fato, pode bem parecer retrógrado fazer o contrário. Mas, dado que modelos diferentes produzem veredictos causais diferentes e dado que não existe uma noção única de *modelo* canônico para uma dada situação (pelo menos nenhuma ainda desenvolvida). Pode parecer que a única opção remanescente é relativizar as relações causais aos modelos (Halpern e Pearl 2005, p. 85)

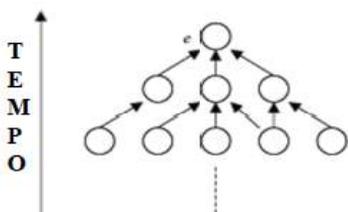
Há pelo menos três respostas ao argumento da modelagem causal a ser consideradas. Primeiro, pode-se evidenciar um ceticismo quanto à modelagem causal. A modelagem causal certamente fornece uma excelente abordagem da epistemologia causal, mas não é óbvio que se deva extrair conclusões metafísicas disso. Segundo, pode-se tentar definir a noção de um modelo canônico. Terceiro, pode-se fornecer uma explicação da causação envolvendo a quantificação existencial sobre modelos (Hitchcock, 2001) ou mesmo a quantificação universal, de modo a utilizar técnicas de modelagem sem pagar qualquer preço adicional na objetividade.

Casos problemáticos: Talvez o principal argumento pela descrição e pela relatividade padrão, que também se apresenta como um argumento pela relatividade do modelo é a utilidade de tal relato adicional na resolução de casos problemáticos. Esse estilo de argumento abrange uma ampla variedade de asserções. Apenas para citar alguns exemplos, McDermott (1995) insere a descrição relativa em uma abordagem geral da causação que goza de uma quantidade razoável de sucesso. McGrath (2005) sugeriu que nossas intuições sobre quais ausências são causas podem vir a ativar expectativas quanto ao que é normal. Hall (2007) mostrou que certos casos causalmente diferentes podem levar a *modelos causais isomórficos*, de modo que técnicas de modelagem possam exigir algo como uma distinção padrão / desviante para discernir esses casos. Hitchcock (2007a) usa modelos com estados padrão atribuídos para definir a noção de *rede causal autocontida*, que ele coloca para trabalhar na abordagem de muitos dos casos mais difíceis na literatura. (Veja Halpern e Hitchcock 2010 bem como Blanchard e Schaffer, para futuros debates sobre a necessidade de uma distinção padrão / desviante dentro da modelagem causal.) É difícil dizer muito mais sobre esses argumentos em geral, sem investigar todos os detalhes. Obviamente, considerações adicionais sobre esses argumentos dependem de considerações detalhadas sobre como as teorias em questão se relacionam com suas rivais, e também sobre o delicado equilíbrio de se a solução ofertada vale o preço da relatividade adicional.

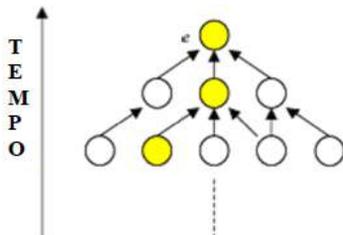
2. Relação

Questão: o que é a relação causal? Quando a bola branca encaçapa a bola nove, qual é a base para esse elo causal? Uma abordagem adequada da relação causal deve revelar por onde as linhas causais correm, que direção às setas causais apontam e se alguma coisa distingue as causas das meras condições de fundo. Em suma, deve revelar a base para a *conexão*, *direção* e *seleção*.

Modelo de rede: A relação causal é tipicamente entendida com referência ao que Steward (1997; veja também Beebee 2004) chama de “o modelo de rede”. O modelo de rede tem dois componentes principais. Primeiro, retrata a relação causal como um segmento direcionado e os relata causais como nódulos. Em segundo lugar, o modelo de rede retrata a história como uma vasta rede causal. No modelo de rede, dadas algumas suposições realistas, a história causal de *e* forma uma vasta árvore invertida (embora uma que, de modo ultimo vai se reduzindo até o big bang):



Uma abordagem da conexão é uma abordagem dos segmentos; uma abordagem de direção é uma abordagem das setas. A seleção no modelo de rede consistiria em destacar certos nódulos:



O modelo de rede sem seleção é implementado, por exemplo, nos diagramas de neurônios popularizados por Lewis. Nos diagramas de neurônios,

os círculos representam duplamente os neurônios que disparam e os eventos que ocorrem, e as setas representam duplamente as sinapses que estimulam e a causação que se dá.

O modelo de rede sem seleção também é implementado nos gráficos acíclicos direcionados usados, na modelagem causal, para visualizar parcialmente os modelos (Spirtes, Glymour e Scheines 1993; Pearl 2000). Em gráficos acíclicos direcionados, os nódulos representam duplamente variáveis com uma gama de valores e ocorrências com uma variedade de alternativas, e as ligações duplamente representam a determinação funcional através das equações estruturais e da paternidade causal. (Para maiores explicações sobre a modelagem causal, veja a entrada: causação: probabilística.)

(Os diagramas de neurônios e os gráficos acíclicos dirigidos são implementações diferentes do modelo de rede. Os neurônios representam eventos que podem ocorrer ou não: eles são de dois valores. Os nódulos representam variável que podem levar dois, três ou mesmo um contínuo de valores. Sinapses representam uma causação atual. Ligações representam linhas de possíveis influências causais, e não implicam qualquer causação real entre os valores reais de suas variáveis. Veja Hitchcock 2007b para comparações e razões para favorecer a abordagem da modelagem causal.)

Conexão: Qual é a base metafísica para a conexão causal? Ou seja, qual é a diferença entre sequências causalmente relacionadas e causalmente não relacionadas?

A questão sobre a conexão ocupa a maior parte da vasta literatura sobre causação. Encontram-se análises de causação em termos de *subsunção nomológica* (Davidson 1980d, Kim 1973, Horwich 1987, Armstrong 1999), *correlação estatística* (Good 1961 e 1962, Suppes 1970, Spirtes, Glymour e Scheines 1993, Kvart 1997 e 2004, Pearl 2000, Hitchcock 2001), *dependência contrafactual* (Lewis 1986a e 2000, Swain 1978, Menzies 1989b, McDermott 1995 e 2002, Ganeri, Noordhof e Ramachandran 1996, Yablo 2002, Sartorio 2005), *manipulação binária agencial* (Collingwood 1940, Gasking 1955, von Wright 1975, Price e Menzies 1993, Woodward 2003), *mudança contígua* (Ducasse 1926), *fluxo de energia* (Fair 1979, Castaneda 1984), *processos físicos* (Russell 1948, Salmon 1984 e 1998, Dowe 1992 e 2000) e *transferência de propriedade* (Aronson 1971, Ehring 1997, Kistler 1998). Também se encontram visões híbridas com relação a alguns dos itens acima (Fair 1979, Dowe 2000, Paul 2000, Schaffer 2001, Hall 2004, Beebe 2004b), juntamente com *primitivismo* (Anscombe 1975, Tooley 1987 e 2004, Carroll 1994, Menzies 1996)

e até mesmo o *eliminativismo* (Russell 1992, Quine 1966).

Felizmente, os detalhes dessas muitas e variadas abordagens podem ser adiados aqui, já que tendem a ser variações de dois temas básicos. Na prática, as abordagens nomológicas, estatísticas, contrafactuais e agenciais tendem a convergir no caso indeterminístico. Todos entendem conexão em termos de probabilidade: causar é tornar mais provável. As abordagens da mudança, da energia, do processo e da transferência convergem no tratamento da conexão em termos de processo: causar é uma produção física. Assim, uma grande parte da controvérsia sobre a conexão pode, na prática, ser reduzida à questão de saber se a conexão é uma questão de probabilidade ou de processo (Seção 2.1).

Direção: Qual é a base metafísica para a direção causal? Ou seja, qual é a diferença entre sequências relacionadas da causa para efeito e aquelas relacionadas do efeito para causa ou do efeito para efeito conjunto de uma causa comum?

A visão padrão sobre a direção da causação é que ela se reduz à direção do tempo: as causas ocorrem antes de seus efeitos (Hume 1975, Kant 1965, Beauchamp e Rosenberg 1981). A visão temporal caiu em desgraça ultimamente, e várias alternativas foram sugeridas, por exemplo, que a direção causal é a direção de *bifurcação* (Reichenbach 1956, Horwich 1987, Papineau 1993, Dowe 2000), *sobredeterminação* (Lewis 1979), *independência* (Hausman 1998) e *manipulação* (Von Wright 1975, Price 1991 e 1996, Woodward 2003). Nessas visões alternativas, a coincidência das ordens temporais e causais é meramente uma característica contingente do mundo atual, ou pelo menos uma característica típica de nossa apreensão dele. Então surge a questão de saber se a direção causal é a direção temporal, ou alguma outra coisa (Seção 2.2).

Seleção: Qual é a base metafísica para a seleção causal? Isto é, qual é a diferença entre sequências de causa e efeitos envolvendo causas reais e aquelas envolvendo meras condições de fundo?

A visão padrão com relação à seleção é que não há *base objetiva* para a seleção: a seleção é orientada por interesse, pragmática e não sistemática (Mill 1846, Lewis 1986a, Mackie 1974). Essa visão sem base objetiva é ocasionalmente desafiada, e alternativas têm sido sugeridas como *suficiente* versus *necessário* (Ducasse 1926) e *anormal* versus *normal* (Hart e Honore 1985). Surge, então, a questão sobre saber se existe alguma diferença objetiva entre “condições desencadeadoras” e “meras condições de fundo” (Seção 2.3).

Pressuposições: A disputa sobre a conexão causal envolve uma pressuposição de *singularidade*; tem que haver uma relação metafísica única de *conectividade causal* para que haja uma questão sobre a natureza dessa relação. No entanto, pode-se argumentar que a linguagem comum permite uma série de noções causais irreduzíveis, como “raspar”, “empurrar”, “molhar”, “carregar”, “comer”, etc. (Anscombe, 1975, p. 68) Por que não usar a linguagem comum do modo como ela se apresenta e adotar uma forma de pluralismo causal em que a "causa" é vista como um agrupamento meramente nominal de relações metafisicamente diversas?

Pode-se defender a singularidade argumentando que raspar, empurrar, molhar e os outros têm uma unidade real (mais do que nominal). Eles têm implicações estatísticas, contrafactuais, preditivas, explicativas e morais comuns. Ou pode-se defender a singularidade como um ideal teórico. Mesmo que o nosso conceito real de causação seja, em última análise, o que Skyrms (1984) chama de “amontoado amável” de princípios (veja também Sober, 1985 e Hall, 2004), isso pode ser considerado um apelo à revisão conceitual. Por que chafurdar na desordem, em vez de arrumar tudo?

2.1. Conexão

Questão: Qual é a base para a conexão causal? É uma questão de probabilidade, processo ou algum híbrido disso? A conexão causal é primitiva e irreduzível? Ou a noção como um todo seria apenas um mito popular?

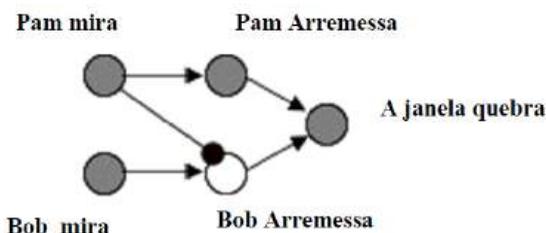
As visões sobre probabilidade e sobre processo aparecem sujeitas a vários problemas sistemáticos, incluindo os problemas de *preempção* e do *fracasso* para visões probabilísticas, e *desconexão* e *má conexão* para visões processuais (Seção 2.1.1). Esses problemas forçam os pontos de vista a evoluir e podem ser vistos como motivadores das alternativas híbridas, primitivista e eliminativista (Seção 2.1.2).

2.1.1. Probabilidade, Processo e Problemas.

Preempção: Um caso problemático para a visão probabilística, no qual a ideia raiz de causação é a de tornar mais provável, é o caso da *preempção* (Good 1961 e 1962, Lewis 1986a, Menzies 1989b, Collins, Hall e Paul 2004, Paul and Hall 2013 *inter alia*). Suponha que Pam e Bob mirem um tijolo em uma janela. Pam joga e quebra a janela, enquanto Bob se contém ao ver Pam em ação. Parece que o arremesso de Pam causou a quebra da janela - o tijolo dela

é o que quebra o vidro. Mas não é necessário que o arremesso de Pam aumente a probabilidade da quebra - se Bob é um vândalo mais eficiente, então o lance de Pam poderia até mesmo tornar menos provável a quebra. Assim, parece que o aumento de probabilidade não é necessário para a causação.

Esse tipo de preempção pode ser representado pelo seguinte diagrama de neurônio (Paul e Hall2013):



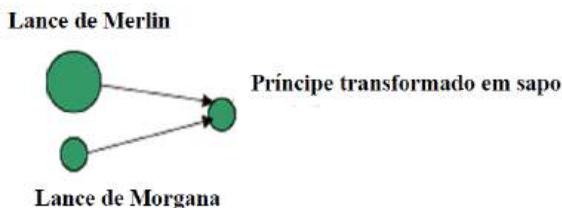
Um círculo preenchido representa duplamente um disparo de neurônio e um evento ocorrendo. A linha com uma seta circular representa duplamente uma conexão inibitória e uma prevenção.

Os teóricos da probabilidade têm respondido à preempção com três estratégias principais, das quais a primeira é *fatorar os intermediários* (Lewis 1986a, Menzies 1989, Eells 1991, Ramachandran 1997, Kvart 1997, Noordhof 1999, Pearl 2000, Yablo 2002). Esta abordagem olha para o curso dos eventos intermediários ou das evoluções de probabilidade, para encontrar algum sentido no qual a causa preventiva é de fato um criador de probabilidade. Por exemplo, se alguém mantiver fixo o fato de que Bob contém seu arremesso, então pode parecer que o arremesso de Pam se torna um arremesso de probabilidade afinal. A segunda resposta que os teóricos da probabilidade oferecem à preempção é *exigir precisão* (Rosen 1978, Lewis 1986a, Paul 2000, Coady 2004). Essa abordagem analisa a maneira e o tempo exato dos eventos envolvidos para tentar encontrar diferenças devido à causa preventiva. Por exemplo, se considerarmos a maneira e o tempo exato de quebra da janela, pode parecer que o lançamento de Pam foi um aumento de probabilidade o tempo todo. A terceira resposta é *considerar a estrutura intrínseca* (Lewis 1986a, Menzies 1996, Hall 2004). Essa abordagem procura a estrutura intrínseca do processo preemptivo, para tentar encontrar algum sentido no qual o processo preemptivo - mesmo que não envolva o aumento de probabilidade - é pelo menos intrinsecamente correto para o aumento de probabilidades.

Não está claro o quão longe essas três estratégias se vão. Existe uma vasta literatura sobre preempção envolvendo uma variedade estonteante de contra-exemplos variados e análises revisadas. Veja Paul e Hall para um excelente guia sobre “tais exemplos grosseiros” (2013, p. 5). É justo dizer que atualmente nenhuma abordagem disponível é capaz de lidar com todos os casos.

Para um caso particularmente difícil, tem sido argumentado que pode haver preempção sem intermediários para fatorar, sem diferenças sobre as quais precisamos de uma melhor precisão ou sem falha intrínseca na preempção da copia de segurança, em casos de *trunfo* (Schaffer 2000a, Lewis 2000). Suponha que as leis da magia digam que o primeiro feitiço lançado em um determinado dia corresponde ao encantamento lançado à meia-noite. Merlin lança um feitiço (o primeiro naquele dia) para transformar o príncipe em um sapo, Morgana lança um feitiço (o segundo naquele dia) para transformar o príncipe em um sapo, e à meia-noite o príncipe se transforma em um sapo. Parece que o feitiço de Merlin fez com que o príncipe se transformasse em um sapo - seu feitiço foi o primeiro lançado naquele dia, e é isso que as leis da magia identificam como a característica relevante. Apenas o feitiço de Merlin satisfaz as condições antecedentes da lei imaginada. Mas não é necessário que o feitiço de Merlin tenha aumentado a probabilidade da transformação - se Morgana é a feiticeira mais confiável, então a chance de transformação teria sido maior se Merlin deixasse o trabalho para Morgana. Não é necessário que haja algum evento intermediário na história - a magia também pode funcionar diretamente. E não parece ser o caso de que haveria diferenças no que aconteceu ao príncipe se Merlin tivesse deixado para Morgana. E, finalmente, não parece haver nenhuma diferença na “aptidão intrínseca” entre o feitiço de Merlin e Morgana. De fato, a única diferença relevante entre esses feitiços parece ser a questão extrínseca de qual foi lançado primeiro.

Esse tipo de superação por preempção pode ser representado pelo seguinte diagrama de neurônio (Paul e Hall 2013):



Aqui nós supomos que os neurônios podem disparar em várias cores (representando várias magias possíveis), e que pela lei, quando um neurônio recebe múltiplos estímulos, ele dispara na cor que combina com a do maior neurônio (que representa o primeiro feitiço naquele dia) o estimulando.

Uma terceira resposta que alguns teóricos da probabilidade têm defendido é de fato aceitar a consequência. Aqui, pode-se afirmar que o efeito ocorreu não por causa, mas *apesar* do preemptor (Eells, 1991, Mellor, 1995). Ou pode-se afirmar que, pelo menos em certos casos, a “preempção do backup” é, na verdade, uma causa que sobredetermina. Por exemplo, Hitchcock (2011) argumenta que uma abordagem contrastiva da causação permite capturar grande parte da assimetria intuitiva dos casos de trunfo, enquanto continua contando o evento superado (a magia de Morgana) como uma causa sobredeterminante do resultado (o príncipe se tornar um sapo).

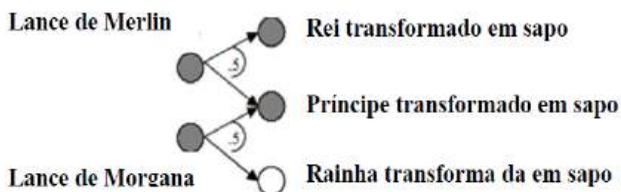
Fracasso: Um segundo caso problemático para a visão da probabilidade é o *fracasso*. Suponha que Pam e Fred apontam um tijolo para uma janela. Pam arremessa e quebra a janela, enquanto Fred simplesmente se afasta, se abaixa ou é impedido por Pam. Parece que a pontaria de Fred não fez a janela se quebrar – pois o tijolo de Fred nunca tocou no vidro. Mas poderia ser o caso de que a pontaria de Fred aumentasse a probabilidade da quebra da janela - se houvesse alguma chance diferente de zero de que Fred fosse bem-sucedido e alguma chance igual a não 1 de que Pam tivesse sucesso, então a pontaria de Fred poderia muito bem ter colocado a janela em um perigo maior. Portanto, parece que o aumento de probabilidade não é suficiente para a causação

A versão de fracasso em que Fred é preemptor pode, é claro, ser representada pelo diagrama dos neurônios do corte inicial acima, com Fred substituindo Bob.

Os teóricos da probabilidade reagiram ao fracasso com as mesmas estratégias que para a preempção, a saber, fatorando intermediários, exigindo precisão ou olhando para a aptidão intrínseca (Menzies, 1989b, Hitchcock, 2004, Kwart, 2004). Por exemplo, se alguém mantiver fixo o fato de que Fred arremessa bem, então pode parecer que o lançamento de Fred não é um aumentador de probabilidade, afinal. Ou, se considerarmos o tempo e a maneira exata de quebra da janela, pode parecer que o arremesso de Fred não foi um aumentador de probabilidade afinal de contas. Ou, se observarmos o caráter intrínseco do processo de Fred, pode parecer que esse processo não estava intrinsecamente apto a causar a quebra da janela.

Não está claro até que ponto essas três estratégias se estendem. Poder-se-ia pensar que pode haver fracassos sem intermediários para fatorar, sem diferenças a fim de ser mais preciso ou sem deficiência intrinsecamente fracassada de não-causa, em casos de *sobreposições* (Schaffer, 2000b). Suponha que Merlin lance um feitiço com uma chance de .5 de transformar o rei e o príncipe em sapos; Morgana lança um feitiço com uma chance independente de .5 de transformar o príncipe e a rainha em sapos; o rei e o príncipe, mas não a rainha transformam-se em sapos. Parece que o feitiço de Morgana não fez com que o príncipe se transformasse em sapo - o fato da rainha não ter sido afetada mostra que o feitiço de Morgana fracassou. Mas é o caso que o feitiço de Morgana aumentou a probabilidade do príncipe se transformar em um sapo. Além disso, não é necessário que haja algum evento intermediário na história (a magia também pode funcionar diretamente). Não é necessário que haja diferenças no que aconteceria ao príncipe, se o feitiço de Morgana tivesse sido adotado em vez do feitiço de Merlin. E não parece ser o caso que o feitiço de Morgana fosse intrinsecamente defeituoso de qualquer forma (na verdade, é apenas a circunstância externa da rainha não ser afetada que revela que o feitiço de Morgana fracassou).

Esse tipo de caso sobreposto pode ser representado pelo seguinte diagrama de neurônio:



Aqui um arco com um número representa um efeito conjuntivo com certa probabilidade.

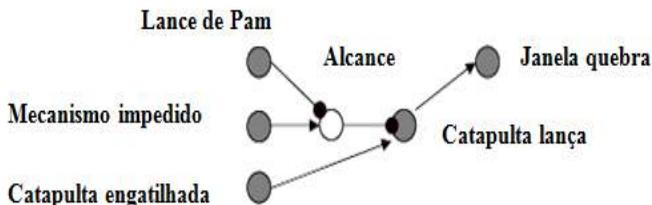
Casos de preempção e fracasso juntos podem ser tomados para mostrar que o aumento de probabilidade (embora interpretado e refinado) é sistematicamente incapaz de prover condições necessárias ou suficientes para a causação. Além disso, esses casos podem sugerir que a conexão é uma questão de processos (sejam eles físicos ou mágicos) ao invés de probabilidades. A causa preventiva e o efeito são ligados por um processo,

enquanto a não causa e o efeito não são - basta olhar para os diagramas. Como Armstrong escreve: “Onde há uma seta em um diagrama mostrando que um neurônio faz com que outro neurônio seja disparado, ou seja, incapaz de disparar, considere que aqui existe uma relação binária genuína de dois termos de causação singular entre a causa e efeito. Onde não há tal seta, negue que haja tal relação. Esta é a porta aberta” (2004, p. 446).

Mas a visão processual enfrenta casos problemáticos.

Desconexão: Um caso problemático para a visão processual, no qual a ideia raiz da causação é a conexão física, é a *desconexão* (Ehring 1984, Schaffer 2000c, Lewis 2004, Hall 2004). Suponha que Pam catapulte seu tijolo pela janela em vez de jogá-lo. Então, parece que Pam catapultando o tijolo causa a janela quebrada - pode realmente importar aqui se Pam catapulta o tijolo ou o joga? Não precisa haver nenhum processo conectando Pam soltando a alavanca e o vôo do tijolo através da janela - nenhum fluxo de momento de energia relevante, faixa de transmissão de marca ou tropo persistente conecta eles. Em vez disso, o que está acontecendo aqui é que a catapulta engatilhada significa impedir que a catapulta lance, e o fato de Pam liberar a alavanca significa impedir que o mecanismo seja impedido de lançar - a catapulta é então liberada. O processo de lançamento é puramente interno à catapulta.

Esse tipo de caso de desconexão pode ser representado pelo seguinte diagrama de neurônio:



A principal resposta feita às desconexões é negar que elas sejam genuinamente causais. Nesse sentido, Aronson diz: “Considere um peso preso a uma mola tensa. Em um determinado momento, a presilha que mantém a mola esticada é liberada e o peso imediatamente começa a acelerar. Pode-se ficar tentado a dizer que a liberação do mecanismo foi à causa da aceleração do peso. Em caso afirmativo, então o que significa a liberação do mecanismo para o peso? Nada, é claro.” (1971, p. 425; ver também Dowe 2001, Hall 2004) De fato,

uma vez que as desconexões envolvem a prevenção de pretensos preventores, e como a prevenção parece envolver ausências, pode-se rejeitar a causalização por desconexão. Pela razão geral de que não há causalização de ausência (Moore 2009). Por exemplo, pode-se negar que qualquer conexão causal possa ocorrer através da ausência do mecanismo.

Má conexão: Um segundo caso problemático para a visão de processo é uma má conexão (Hitchcock 1995b, Dowe 2000, Schaffer 2001). Suponha que Pam atire seu tijolo na janela, enquanto o inocente Tom assiste consternado, ou borrife tinta roxa no ar através do qual o tijolo de Pam passa. Então parece que a observação de Tom ou a pulverização de tinta não causam a quebra da janela. Mas há um processo que liga a observação de Tom ou a pulverização de tinta à ruptura. Quando Tom assiste, haverá fótons ligando-o à quebra. Quando Tom borrifar tinta no tijolo, haverá uma faixa de tinta roxa da lata de spray no tijolo até a janela. (As más conexões podem ser subdivididas em micro-conexões, que são de magnitude inadequada, como os fótons e as pseudo-conexões, que são do tipo incorreto, como a pintura.) Encontramos duas principais respostas para as conexões inadequadas na literatura, a primeira é de fato encarar as consequências. No caso da conexão entre os fótons da observação de Tom e a quebra da janela (e micro-conexões em geral), isso pode ser considerado como causalização de proporções tão insignificantes que é compreensível que possamos negligenciá-la. Nossas intuições em contrário também podem ser descartadas, no estilo de Davidson, como confundir idiomas causais e explicativos.

A segunda resposta principal encontrada para as conexões inadequadas é refinar os processos envolvidos. No caso da conexão entre a pulverização de tinta de Tom à quebra da janela, a linha de persistência da tinta e a linha do tijolo que voa pela janela podem ser consideradas distintas e meramente coincidentes (Dowe, 2000). Desta forma, pode-se negar que existe um processo genuíno conectando a não-causa de conexão inadequada ao efeito.

Casos de desconexão e má conexão juntos podem ser tomados para mostrar que a conexão entre processos (embora interpretada e refinada) é sistematicamente incapaz de fornecer condições necessárias ou suficientes para a causalização. Além disso, esses casos podem sugerir que a conexão é antes uma questão de probabilidades do que de processos. Enquanto a causa desconectada (como a liberação da catapulta por Pam) e o efeito estão ligados por probabilidade, a não causa de má conexão (como Tom observando ou

pintando com spray) e o efeito não está. Então, o teórico da probabilidade pode reivindicar vingança. Mas também é possível extrair a moral mais ampla de que as visões de probabilidade e processos são, na melhor das hipóteses, aspectos de um quadro maior. Isso pode inspirar a busca por uma visão híbrida.

2.1.2. Híbridos, Primitivos e Eliminativos

Híbridos: Dada a plausibilidade intuitiva das visões de probabilidade e processo, e os problemas sistemáticos que cada um encontra, alguns teóricos recentes buscaram uma reconciliação. As visões híbridas visam sintetizar as visões de probabilidade e processo, capturando o que está intuitivamente correto sobre ambas as ideias enquanto resolvem seus casos problemáticos.

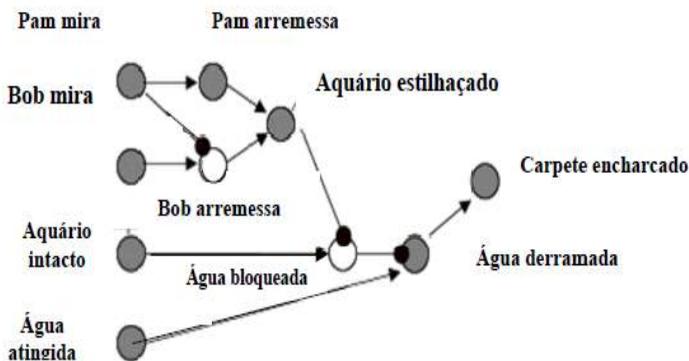
As visões híbridas mais óbvias simplesmente combinam ou desarticulam as visões de probabilidade e processo, ou postulam uma ambiguidade entre as noções (Hall, 2004). Com relação aos quatro casos problemáticos acima, pode parecer pouco promissor. Requerer que ambos, a probabilidade e o processo, resolvam os fracassos e as más conexões (Salmon 1997), uma vez que o primeiro não envolve nenhum processo de ligação e o segundo não envolve nenhum aumento de probabilidade. Mas, pela mesma razão, não resolverá as preempções e desconexões, uma vez que o primeiro não envolve o aumento de probabilidade e o segundo não envolve conexões entre processos. (O teórico conjuntivo pode ainda se valer de uma das respostas acima para a preempção e as desconexões). Da mesma forma, pode parecer ingênuo requerer que a probabilidade ou o processo resolva as preempções e desconexões, mas não os fracassos e as más conexões. (O teórico disjuntivo ainda pode se beneficiar de uma das respostas acima para esses problemas.

As visões híbridas mais sofisticadas tentam *integrar* as noções de probabilidade e de processo, e não apenas unir ou separar as mesmas. Assim, Fair (1979), em última análise, passa de uma visão de fluxo de energia para uma visão que entende a conexão em termos de contrafactuais sobre o fluxo de energia. E Schaffer (2001) sugere uma generalização dessa abordagem, na qual a conexão causal é entendida em termos de probabilidade dos processos. Esse tipo de visão híbrida pode parecer resolver todos os casos de problemáticas acima. Enquanto a causa preventiva e a causa desconexa aumentam a probabilidade do processo que produz o efeito, a não causa fracassada e a não causa mal conectada não aumentam.

No entanto esse tipo de visão híbrida tem dificuldades com casos

problemáticos “hibridizados”, como os apresentados a seguir (Schaffer, 2001). Suponha que Pam arremesse um tijolo na parede vidro de um aquário, impedindo Bob de fazer o mesmo. O aquário estilhaçado causa a imersão do carpete, impedindo que o vidro impeça a água de derramar. Este é um caso de preempção inserido em um caso de desconexão. Parece que o arremesso de Pam causou a imersão do tapete - o tijolo dela é o que quebrou o aquário. Mas não é necessário que o arremesso de Pam aumente a probabilidade do processo de produzir a imersão - se Bob for um vândalo mais eficiente, então o lance de Pam poderia até mesmo diminuir a chance do processo de derramamento, por preempção de Bob. E também pode ser que a mira de Bob tenha aumentado a probabilidade do processo que produziu a imersão - a mira de Bob poderia ter aumentado a chance do processo de derramamento, ao ameaçar o aquário.

(Tal caso também cria problemas para a visão de “dois conceitos” do estilo Hall, uma vez que o lance de Pam acaba por não satisfazer *nenhum* dos dois conceitos de causação relatados por Hall.) Esse tipo de caso de desconexão preventiva pode ser representado pelo seguinte diagrama dos neurônios:



A literatura atual sobre causação está agora repleta de casos híbridos complexos, incluindo casos de preempção preemptiva (McDermott 1995, Collins 2000), desconexões dentro de cadeias maiores (Hall 2004) e uma variedade de variações de preempção desviantes (Paul e Hall 2013). Nenhuma abordagem conhecida de causação - híbrida ou não - dá conta de todos esses casos. Portanto, não está claro o que as abordagens híbridas ganham no final.

Primitivismo: Os problemas encontrados pelas tentativas de analisar a

conexão causal fornecem um dos argumentos principais para o primitivismo. Parece haver um padrão de falha, que pode sugerir que a causalção é simplesmente não analisável.

O segundo argumento principal para o primitivismo é que a causalção é *central demais* para ser reduzida. As abordagens de probabilidade e processo (e por extensão, híbridos) são inevitavelmente circulares, porque as noções de probabilidade e de processo não podem ser entendidas sem referência à causalção. Quanto à probabilidade, pode-se pensar que cada uma das versões nomológica, estatística, contrafactual e agencial da teoria abrigam noções causais. A relação causal poderia ser necessária para distinguir leis reais de generalizações acidentais (Armstrong 1997), para distinguir quais condições de fundo devem ser mantidas fixas na avaliação estatística das quais podem variar (Cartwright 1983), para distinguir quais condições de fundo podem ser mantidas fixas em suposições contrafactuais e quais podem variar (Kvart 1986), e para compreender a noção de intervenção agencial (Hausman 1998). Quanto ao processo, pode-se pensar que um processo nada mais é do que uma sequência causal; nas palavras de Sayre: "O processo causal, por mais contínuo que seja, é constituído de eventos individuais relacionados a outros em umnexo causal... É em virtude das relações entre os membros da série causal que somos capazes de fazer as inferências pelas quais os processos causais são caracterizados." (1977, p. 206)

De fato, o primitivista poderia acrescentar que a noção de evento (ou qualquer que sejam os relata causais que tomemos) não pode ser entendida sem referência à causalção, porque as próprias propriedades são individuadas por seus papéis causais (Shoemaker 1980 e 1998, Ellis 1999). Como Carroll observa a esse respeito, "com relação ao nosso aparato conceitual total, a causalção está no centro do centro". (1994, p. 118) Assim, poderíamos pensar que a análise é impossível, porque não temos conceitos mais básicos.

Como um terceiro argumento principal para o primitivismo, tem sido argumentado que existem mundos que diferem de forma puramente causal (Armstrong 1983, Tooley 1987 e Carroll 1994). Suponha que seja uma lei da magia que os feitiços lançados por Merlin e Morgana tenham uma chance de .5 de transformar o príncipe em sapo à meia-noite e que o príncipe se transforma em sapo à meia-noite. Então, pode ser intuído que existam três possibilidades distintas: uma em que apenas o feitiço de Merlin causou a transformação, um segundo no qual apenas o feitiço de Morgana o fez, e um terceiro em que ambos o fizeram. Essas possibilidades apresentam as mesmas leis e padrões de

eventos. Assim, alguém pode alegar intuir que a causação é ontologicamente fundamental.

Mas há também três argumentos principais contra o primitivismo, o primeiro remonta a Hume e sustenta que o primitivismo entra em conflito com a existência do conhecimento causal. Afinal, pode-se argumentar (em uma veia humiana) que tudo o que podemos observar são sequências de eventos; assim sendo, nunca poderíamos conhecer quaisquer fatos sobre a conexão causal, se a conexão é algo acima e além de tais sequências. A isto, o primitivista poderia responder, ou que as conexões primitivas podem ser diretamente observadas, pelo menos em certos casos favoráveis, tais como o desejo ou a pressão sobre o corpo (Anscombe 1975, Strawson 1985, Fales 1990, Armstrong 1997); ou que as conexões primitivas podem ser inferidas teoricamente por meio de uma inferência da melhor explicação (Tooley, 1987). O segundo argumento principal contra o primitivismo é que a causação primitiva é um tipo assustador de modalidade primitiva. Na medida em que as relações causais supostamente são *conexões necessárias* na natureza, e na medida em que geralmente favorecemos a redução do modal à ocorrência, teremos uma razão geral para resistir a qualquer primitivismo causal.

O terceiro argumento principal contra o primitivismo é que ele leva ao eliminativismo. Pois se as opções são causação irreduzível ou nenhuma, deve-se perguntar se "nenhuma" pode ser a melhor escolha. Pois, se a ciência fornece o critério para quais entidades contingentes básicas devemos reconhecer então que a questão que deveria surgir é se alguém poderia fazer ciência sem qualquer primitivismo causal que seja. Pode-se concluir que nossa noção popular de causação deve ser reduzida ou enfrentar o eliminativismo.

Eliminativismo: A visão final sobre a conexão causal a ser considerada é a visão eliminativista, conforme proclamada por Russell: "A lei da causação... é uma relíquia de uma era passada, sobrevivendo, como a monarquia, somente porque erroneamente se supõe que não cause dano algum. (1992, p. 193) O eliminativista vê o conceito causal como uma projeção ingênua da agência sobre o mundo, a ser superada em um esquema científico sofisticado.

O principal argumento para o eliminativismo é que a ciência não precisa de causa. A noção de causação é vista como uma relíquia cientificamente retrógrada da metafísica da Idade da Pedra. Como Russell afirma: "Nos movimentos dos corpos que se gravitam mutuamente, não há nada que possa ser chamado de causa e nada que possa ser chamado de efeito; existe apenas uma fórmula." (1992, p. 202, ver também Quine 1966) Diz-se que as equações

diferenciais da física sofisticada não deixam espaço para causas, ou pelo menos não precisam delas.

O argumento de Russell pode ser eficaz contra um tratamento primitivista da causação, mas o reducionista pode responder que a causação ainda se reduz a entidades cientificamente respeitáveis. A esse respeito, "evento", "lei", "causa" e "explicação" estão no mesmo barco. Esses termos nômicos servem para permitir uma compreensão sistemática da ciência; eles não aparecem nas equações. A partir dessa perspectiva, o argumento de Russell pode parecer semelhante à afirmação tola de que a matemática eliminou a variável, porque o termo "variável" não aparece nas equações!

A principal objeção ao eliminativismo é que a causação é central demais para ser eliminada. A causação, segundo vários filósofos contemporâneos é necessária para a análise de conceitos metafísicos como persistência, conceitos científicos como explicação e disposição, conceitos epistêmicos como percepção e garantia, conceitos éticos como ação e responsabilidade, conceitos mentais como papel funcional e conteúdo conceitual e conceitos linguísticos como referência. O eliminativismo não é apenas injustificado; mas também catastrófico. Assim, pode parecer que os argumentos contra o primitivismo e contra o eliminativismo trazem de volta uma abordagem reducionista da causação e, portanto, levam de volta à probabilidade, ao processo, e aos seus descendentes.

Embora talvez exista algum meio termo entre a tarefa sem esperança da análise conceitual, por um lado, há o desastre conceitual do primitivismo/eliminativismo. E por outro não se deve confundir as perspectivas de análise conceitual com as perspectivas de redução ontológica. Pode bem ser que o nosso conceito de causação - algo na mente - não seja definível através de outros conceitos (de um modo que contaria como uma análise conceitual informativa). Contudo, também pode ser que a relação causal - uma entidade fora do mundo - não seja um constituinte fundamental da realidade. Essa posição intermediária pretenderia explicar tanto as falhas das análises conceituais quanto o desaparecimento da causação da física fundamental. (Veja Schaffer 2007, pp. 872-873, para mais alguma discussão.)

2.2. Direção

Questão: Qual é a base para a direção causal? É uma questão de direção temporal, ou algo mais como a direção da bifurcação, a direção da

sobredeterminação, a direção da independência ou a direção da manipulação? O direcionamento causal é um assunto primitivo e irredutível? É o direcionamento causal apenas um mito popular, ou talvez uma projeção de nossa experiência como agentes humanos em um mundo não direcionado?

Na prática, encontramos seis principais argumentos sobre a questão de se devemos identificar direção causal com direção temporal. Primeiro, há o argumento de fraudar, que sustenta que a ordem causal deve ser a ordem temporal, ou então o efeito pode ocorrer, mas a causa é evitada. Segundo, o argumento *da viagem no tempo*, que sustenta que a ordem causal não deve ser a ordem temporal, devido à possibilidade de viagem no tempo. Terceiro, o argumento *da causação simultânea*, que sustenta que a ordem causal não deve ser a ordem temporal devido à possibilidade de a causa e o efeito serem simultâneos. Quarto, o argumento dos *efeitos conjuntos*, que sustenta que não ajudará analisar a ordem causal como ordem temporal, porque há casos de efeitos conjuntos nos quais há uma conexão temporalmente ordenada sem causação. Quinto, o argumento *da física*, que sustenta que a ordem causal não deve ser a ordem temporal por causa de várias hipóteses físicas que envolvem a causação reversa. E sexto, o argumento *da teoria causal do tempo*, que sustenta que a ordem causal é a ordem temporal, mas apenas porque a ordem temporal deve ser analisada em termos da ordem causal, e não *vice-versa*.

Fraudar: O argumento principal para a ordem causal ser a ordem temporal é o argumento de fraudar (Black, 1956). O argumento é explicado por Mackie, com referência a uma hipótese causal retroativa de que um desenho feito por um suposto clarividente na segunda-feira poderia ser causado por um padrão feito na terça-feira: “Mas em todas as ocasiões, depois de feito o desenho, é possível que alguém ou algo intervenha para que o padrão correspondente não seja produzido. Consequentemente, não pode em nenhuma ocasião, ser o padrão responsável pelos detalhes do desenho: a hipótese de precognição deve ser falsa, mesmo para aquelas ocasiões em que o dispositivo não é parado, quando o padrão é realmente produzido e acaba sendo exatamente como o desenho.” (1974, p. 178) Assim, a causação temporalmente reversa é considerada impossível.

Existem duas respostas para o argumento de fraudar. A primeira apresentada por Dummett(1964), consiste em observar que o argumento só se aplica a casos em que a intervenção humana é possível. O que há para evitar a causação reversa quando a intervenção humana é descartada?

A segunda resposta é que o argumento envolve uma mistura incoerente

de determinismo e indeterminismo. Se o mundo é determinista, então a intervenção fraudulenta é impossível, pois já será dado que a causa posterior ocorrerá. Se o mundo é indeterminista, então a intervenção fraudulenta é possível, mas não mais problemática, pois o caso reduzir-se-á àquele em que o evento anterior (por exemplo, o desenho do clarividente) é uma erupção indeterminista sem causa.

Viagem no tempo: O primeiro argumento principal contra a ordem causal ser a ordem temporal é que a causalização temporalmente reversa é possível em casos como a *viagem no tempo*. Parece metafisicamente possível que um viajante do tempo entre em uma máquina do tempo no tempo t_1 , fazendo com que ela saia da máquina do tempo mais cedo t_0 . De fato, isso parece ser possível em termos nomológicos, já que Gödel provou que há soluções para as equações de campo de Einstein que permitem percursos em looping: “Ao fazer uma viagem de ida e volta em um foguete é possível viajar [nos mundos governados pelas equações de campo de Einstein] para qualquer região do passado, presente e futuro e vice-versa, exatamente como é possível, em outros mundos, viajar para partes distantes do espaço.” (1949, p. 560)

Existem três respostas para o argumento da viagem no tempo. A primeira resposta é que a viagem no tempo é *incoerente*. Uma variedade de incoerências pode ser alegada aqui, incluindo a incoerência de mudar o que já está fixo (causando o passado), de ser capaz e incapaz de matar os próprios ancestrais, ou de gerar um ciclo causal e, portanto, uma relação reflexiva de “auto-causação”, ou de gerar atribuições de probabilidade inconsistentes (Mellor 1995). A prova de Gödel pode ser descartada como um mero artefato matemático, não refletindo qualquer situação possível.

A segunda resposta ao argumento da viagem no tempo é que a viagem no tempo ainda pode ocorrer através de etapas causais localmente futuras. De fato, é exatamente isso que acontece nos casos nomologicamente possíveis descobertos por Gödel; O espaço-tempo é estruturado topologicamente de tal maneira que uma série de etapas de avanço local produz um caminho globalmente inverso. Isso é compatível com a ordem causal, sendo a ordem temporal, pelo menos em cada etapa particular. Pode ser que tanto a ordem causal quanto a ordem temporal possam deixar de possuir orientação global.

A terceira resposta ao argumento da viagem no tempo é que qualquer caso alegado de viagem no tempo está aberto para redescritão. Ao invés de o viajante do tempo entrar na máquina em t_1 e sair em t_0 , a mesma situação pode ser redescrita em termos da criação espontânea em t_0 de um indivíduo, e o

desaparecimento espontâneo em t_1 de outro, com correlações meramente coincidentes entre seus vários estados mentais e físicos.

Causação Simultânea: O segundo principal argumento contra a ordem causal ser a ordem temporal, é que a *causação simultânea* é possível. De fato, pode parecer que a causação simultânea ocorre no mundo real, por exemplo, quando uma bola de ferro comprime uma almofada (Kant, 1965, Taylor, 1966, Brand, 1980).

A principal resposta ao argumento da causação simultânea é que os casos que parecem exemplificá-lo são *mal descritos* (Mellor, 1995). A bola de ferro leva tempo para bater na almofada e, em geral, todos os corpos demoram a comunicar seus movimentos. Não há corpos perfeitamente rígidos, pelo menos em qualquer mundo nomologicamente possível. Sem o suporte intuitivo deste tipo de caso, o argumento da causação simultânea pode ser acusado de cometer petição de princípio. Neste ponto, questões metodológicas sobre a relevância de casos concebíveis, mas fisicamente impossíveis, podem surgir.

Física: O terceiro principal argumento contra a ordem causal ser a ordem temporal é o argumento da física. Físicos do século passado entretiveram uma variedade de teorias que postulavam a causação retrógrada, incluindo a teoria de radiação de Wheeler-Feynman, a teoria dos táquions de Feynman e sua teoria dos pósitrons como elétrons voltando no tempo, a explicação do “aperto de mão quântico” de Beauregard e da violação das desigualdades de Bell. Embora nenhuma dessas teorias tenha muito crédito hoje, elas eram hipóteses físicas sérias que pareciam estabelecer que elas ao menos poderiam ter sido verdadeiras (Horwich 1987, Dowe 2000).

Há duas respostas principais ao argumento da física, o primeiro é *dispensar* essas teorias. Talvez essas teorias sejam todas falsas ou mesmo sutilmente incoerentes. Aqui, a possibilidade de redescrição posterior aludida com a viagem no tempo é saliente, na medida em que o defensor da ordem temporal pode sustentar que a redescrição antecipada é sempre possível, e sempre preferível.

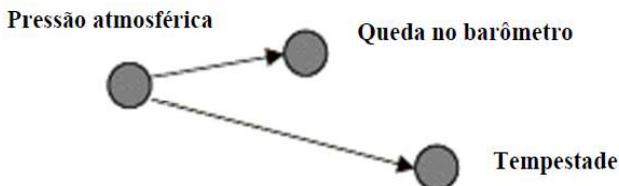
A segunda resposta ao argumento da física é que ele *extrapola*. Pode não haver uma abordagem coerente da ordem causal compatível com todas essas teorias. Em particular, o modelo causal retrógrado da violação das desigualdades de Bell postula uma seta causal para trás que, ao que parece, não é nem o pedaço de uma bifurcação em direção ao passado, muito menos um sobredeterminador especial do futuro, nem um manipulador para manipular o

passado. Assim, o argumento da física pode culminar em um geral *tu quoque*.

Um tipo diferente de argumento da física não olha para as teorias que apontam para a causação reversa, mas sim para a falta de quaisquer assimetrias relevantes (temporais ou não) na física avançada. A partir disso infere-se que a própria realidade pode ser temporalmente simétrica e que qualquer sentido de *direção* para as relações causais se deve a uma projeção de nossa experiência como agentes (Price, 1996 e 2007). Alternativamente, pode-se argumentar que a física fornece uma direção real - talvez através da "Hipótese do Passado" de Albert (Albert 2000, Loewer 2007, Kutach 2007) - mas uma que está apenas contingentemente conectada à direção do tempo.

Efeitos Conjuntos: O terceiro argumento principal contra a ordem causal ser a ordem temporal é o problema dos efeitos conjuntos (Lewis, 1986a). Suponha que a queda na pressão atmosférica em t_0 cause tanto a queda no barômetro em t_1 quanto a tempestade em t_2 . Então a queda no barômetro e a tempestade são causalmente conectadas e temporalmente ordenadas, mas isso não é causação, e sim um caso de efeitos conjuntos de uma causa comum.

Este caso pode ser representado pelo seguinte diagrama de neurônios:



Aqui, a ordem da esquerda para a direita representa a ordem temporal.

Há duas respostas para o argumento dos efeitos conjuntos, a primeira é adicionar alguns outros testes para uma estrutura de efeito conjunto, como o teste de triagem (Reichenbach 1956, Suppes 1970, inter alia). A direção causal é então tomada como a direção temporal não triada.

A segunda resposta é restringir a visão da ordem temporal aos casos de conexão direta. Como o diagrama dos neurônios sugere, os efeitos conjuntos são apenas indiretamente conectados, através de sua causa comum (Horwich, 1987). Se for possível identificar as conexões diretas e aplicar a ordem temporal apenas a elas, teríamos uma seta desde a queda da pressão atmosférica até a queda no barômetro, bem como a tempestade, mas não há nenhuma seta do barômetro à tempestade. Isso corresponderia ao diagrama.

Teoria Causal do Tempo: um argumento final a ser considerado é o argumento de que a ordem temporal deve ser analisada em termos da ordem causal, e não *vice-versa* (Kant, 1965, Reichenbach, 1956, Mellor, 1981). Esse argumento corta para os dois lados, no sentido de que implica que a ordem causal é a ordem temporal (*contra* os argumentos da viagem no tempo, causação simultânea e física), e no sentido em que implica que a ordem causal não pode ser *baseada* na ordem temporal sob a influência da circularidade.

A principal resposta ao argumento da teoria causal do tempo é, obviamente, rejeitar a teoria causal do tempo. A direção temporal terá que ser entendida em outros termos, talvez em termos de assimetrias físicas intrínsecas, tais como as envolvida com a entropia ou com a decadência do kaon neutro, ou talvez simplesmente tomada como primitiva (Maudlin 2007). Embora na medida em que se toma a direção como primitivo, existe a tentação de colocar este primitivo em ação o máximo possível, incluindo a compreensão da direção causal.

2.3. Seleção

Questão: Qual é a base para a seleção causal? A distinção entre causas reais e meras condições de fundo é apenas um assunto arbitrário e não sistemático? Ou existe uma base metafísica para seleção, relacionada à suficiência versus necessidade, a anormalidade versus normalidade ou a qualquer outra coisa?

(Deve-se notar que a seleção está amplamente associada à ideia de "causa". Talvez isso seja um erro. Muitas vezes selecionamos múltiplas causas que agem juntas. Por exemplo, em um caso de *causação conjunta*, quando os quatro carregadores carregam o piano coletivamente pelas escadas, seria natural selecionar os esforços de cada indivíduo como uma causa real do piano atingir o segundo andar (selecionando quatro causas reais), enquanto, rebaixam-se vários fatores, da presença da escadaria até o status das condições de fundo. "A causa", relativa a um determinado contexto, simplesmente se refere a algo como a causa mais saliente no contexto, assim como "o cachorro" se refere a algo como o cão mais saliente no contexto; isso tudo é uma questão sobre o que o "o" significa e nada tem a ver com a causação, ou com os cachorros nesses casos.)

Na prática, podemos encontrar quatro principais argumentos sobre a questão da seleção. Primeiro, o argumento do capricho, que sustenta que nossa

prática real de seleção é muito caprichosa para ter qualquer base real. Segundo, o argumento da previsibilidade, que sustenta que nossa prática real de seleção é previsível demais para não ter uma base real. Terceiro, o argumento da inseparabilidade, que sustenta que não temos nenhum conceito de causação que subtraia a seleção. E quarto, o argumento da aridade revisitada, que alega que os dados adicionais podem reconciliar os três argumentos anteriores.

Capricho: O principal argumento para a visão sem base é o argumento de Mill do capricho: “Nada pode mostrar melhor a ausência de qualquer base científica para a distinção entre a causa de um fenômeno e suas condições, do que a maneira caprichosa em que nós selecionamos dentre as condições aquelas que nós escolhemos para denominarmos a causa” (1846, p. 198). O argumento de Mill ganhou o campo, e é ecoado por autores contemporâneos como Lewis: “Às vezes escolhemos uma entre todas as causas de algum evento e chamamos de "a" causa, como se não houvesse outras. Ou destacamos algumas como "causas", chamando o resto de meros "fatores causais" ou "condições causais"... Podemos selecionar as causas anormais ou extraordinárias, ou aquelas sob controle humano, ou aquelas que consideramos boas ou más, ou apenas aquelas que queremos falar. Não tenho nada a dizer sobre esses princípios de discriminação injusta.” (1986, p. 162) Assim, a seleção é geralmente descartada como infundada, e os teóricos procuram isolar alguma concepção igualitária pré-selecionada de causação.

Previsibilidade: O argumento principal contra a visão sem base sustenta que nossas seleções são muito previsíveis para não ter base. Este ponto foi feito por Hart e Honore, que escrevem: “Na maioria dos casos em que um incêndio aconteceu, o advogado, o historiador, e o homem comum se recusaria a dizer que a causa de o fogo era a presença de oxigênio, embora nenhum fogo tivesse ocorrido sem ele: eles reservariam o título de causa para algo da ordem de um curto-circuito, a queda de um cigarro aceso ou um raio... Ao fazer essa distinção, Está claro que nossa escolha, embora seja responsiva ao contexto variável das ocasiões particulares, não é arbitrária ou aleatória.” (1985, p. 11)

Mas o que poderia ser essa distinção entre causas e condições? Ducasse sustenta que é entre causas suficientes e condições necessárias: “Como uma questão de uso estabelecido, "causa" é contrastada com "condição" de uma maneira útil e claramente declarada: A causa de um fenômeno é uma mudança em suas circunstâncias antecedentes que foi *suficiente* para fazê-lo ocorrer. Uma *condição* de um fenômeno, por outro lado, é uma mudança, ou

mais frequentemente um estado, de suas circunstâncias antecedentes, o que era *necessário* para que ocorresse quando aconteceu.” (1969, p. 19) Mas é difícil enxergar como isso captura nossa seleção do curto-circuito sobre a presença de oxigênio, pois cada fator parece necessário, mas não suficiente.

Hart e Honore sustentam que situações anormais e ações livres são causas, enquanto situações normais e fatores não-agenciais são condições: “Ao distinguir causas e condições, dois contrastes são de fundamental importância. Estes são os contrastes entre o que é anormal e o que é normal em relação a qualquer coisa ou assunto, e entre uma ação deliberadamente livre e todas as outras condições.” (1985, p. 33) Isto parece fazer mais sentido com relação ao curto-circuito (anormal) versus a presença de oxigênio (normal), mas ao preço de tanta vagueza, que se pode pensar que em uma visão seria apenas verbalmente distinto da visão sem base.

Inseparabilidade: Um argumento adicional contra a visão sem base é que não temos nenhum conceito de causação sem seleção. Como Hart e Honore colocam: “O contraste de causa com meras condições é uma característica inseparável de todo pensamento causal, e constitui o significado de expressões causais tal como a referencia implícita a generalização o faz.” (1985, p. 12, ver também Schaffer 2005). O resultado desse argumento é que a visão sem base nos priva de qualquer compreensão intuitiva da noção de causa. Pois como é que vamos julgar se certos casos, como os casos problemáticos revistos acima ou quaisquer outros, envolvem causação ou não, se os nossos julgamentos estão infectados com um componente de capricho não sistemático?

Lewis escreve: “Estou preocupado com a questão prévia do que é ser uma das causas (falar de maneira não-seletiva). Minha análise destina-se a capturar um conceito amplo e não discriminatório de causação” (1986a, p. 162). Mas não é óbvio que *tenhamos* qualquer conceito como o que Lewis procura. Ou, pelo menos, não é óbvio que nossas intuições sobre a causação possam fornecer qualquer *evidência* a respeito desse “conceito amplo e não discriminatório”, se nossas intuições forem afetadas por efeitos de seleção.

Aridade revisitada: Um último argumento a ser levado em conta na questão da seleção revisita a aridade (Seção 1.3), e sustenta que os relatos causais adicionais podem reconciliar o capricho e a previsibilidade e explicar a inseparabilidade (Schaffer, 2005 e 2012). O que é caprichoso sobre a seleção é que falantes diferentes, em diferentes contextos de conversação, discordarão sobre o que é uma causa e o que é uma condição de fundo. Se alguém não

sabe qual pergunta um orador está seguindo, pode-se achar as escolhas dela caprichosas. O que é previsível sobre a seleção é que, uma vez em que o contexto conversacional seja fixo, pode-se esperar um amplo acordo sobre causas versus condições. Se soubermos quais inquéritos um orador está investigando, veremos que suas seleções são previsíveis. Isso pode sugerir o que está variando caprichosamente, será quais contrastes estão em jogo em uma dada investigação e o que é previsível é o que contam como uma causa real *em relação aos contrastes em jogo*.

Mackie (um teórico binário) fala do *campo causal* em referência a qual seleção causal é feita: "Uma declaração causal será a resposta a uma questão causal, e a pergunta "O que causou essa explosão?" Pode ser expandida em "O que fez a diferença entre aqueles tempos, ou aqueles casos, dentro de um certo intervalo, em que tal explosão não ocorreu, e este caso em que ocorreu uma explosão?" Ambas as causas e efeitos são vistos como diferenças dentro de campo; qualquer coisa que seja parte da descrição assumida (mas comumente não declarada) do campo em si será, então, automaticamente descartada como um candidato para o papel da causa." (1974, p. 35) O contrastivista pode oferecer uma implementação natural de noção de Mackie de um campo causal, entendida agora como os aspectos da situação que são assumidos presentes e para os quais alternativas (/ contrastes) não são consideradas. Isso é particularmente apropriado para entender a seleção no contexto dos modelos causais, nos quais apenas certos eventos são representados por meio de variáveis enquanto (eventos que são representados por variáveis são equipados com uma gama de valores alternativos; eventos que não são representados por variáveis não são).

A seleção causal, em todas as visões, é um reflexo de quais alternativas estão em questão. Para o teórico quaternário, a seleção é determinada pelos valores dos relata contrastante (incluindo a questão de se qualquer contraste que seja está sendo considerado para um dado fator). Ela é, portanto, um componente inseparável do nosso conceito causal. Para a pessoa que pensa que a causação é relativa a um modelo causal, a seleção pode ser determinada de forma comparável a gama de eventos modelados por variáveis (que são equipados com uma variedade de alternativas). Nessas visões, a própria noção de conexão causal é apenas bem definida à luz de contrastes e / ou modelos, e esses relata adicionais explicam a seleção.

Referências

- Achinstein, Peter, 1975, "Causation, Transparency, and Emphasis," *Canadian Journal of Philosophy*, 5: 1–23.
- , 1983, *The Nature of Explanation*, Oxford: Oxford University Press.
- Albert, David, 2000, *Time and Chance*, Cambridge, MA: Harvard University Press.
- Anscombe, G. E. M., 1975, "Causality and Determination," em E. Sosa (ed.), *Causation and Conditionals*, Oxford: Oxford University Press, pp. 63–81.
- Armstrong, D. M., 1997, *A World of States of Affairs*, Cambridge: Cambridge University Press.
- , 1999, "The Open Door," em H. Sankey (ed.), *Causation and Laws of Nature*, Dordrecht: Kluwer Academic Publishers, pp. 175–85.
- , 2004, "Going through the Open Door Again: Counterfactual versus Singularist Theories of Causation," em J. Collins, N. Hall, e L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 445–57.
- Aronson, Jerrold, 1971, "On the Grammar of "Cause"," *Synthese*, 22: 414–30.
- Beauchamp, Tom e Alex Rosenberg, 1981, *Hume and the Problem of Causation*, Oxford: Oxford University Press.
- Beebee, Helen, 2004a, "Causing and Nothingness," em J. Collins, N. Hall, e L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 291–308.
- , 2004b, "Chance-changing Causal Processes," em P. Dowe e P. Noordhof (eds.), *Cause and Chance: Causation in an Indeterministic World*, London: Routledge, pp. 39–57.
- Bennett, Jonathan, 1988, *Events and their Names*, Indianapolis: Hackett Publishers.
- Black, Max, 1956, "Why Cannot an Effect Precede its Cause," *Analysis*, 16: 49–58.
- Blanchard, Thomas e Jonathan Schaffer, forthcoming, "Cause without Default," em H. Beebee, C. Hitchcock, e H. Price (eds.), *Making a Difference*, Oxford: Oxford University Press.
- Brand, Myles, 1980, "Simultaneous Causation," em P. van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*, Dordrecht, D. Reidel Publishing, pp. 137–53.

- Campbell, Keith, 1990, *Abstract Particulars*, Oxford: Basil Blackwell.
- Carroll, John, 1994, *Laws of Nature*, Cambridge: Cambridge University Press.
- Cartwright, Nancy, 1983, *How the Laws of Physics Lie*, Oxford: Clarendon Press.
- Castañeda, Hector-Neri, 1984, "Causes, Causity, e Energy," em P. French, T. Uehling, Jr., e H. Wettstein (eds.), *Midwest Studies in Philosophy* (Volume IX), Minneapolis: University of Minnesota Press, pp. 17–27.
- Coady, David, 2004, "Preempting Preemption," em J. Collins, N. Hall, e L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 325–40.
- Collingwood, R. G., 1940, *An Essay on Metaphysics*, Oxford: Clarendon Press.
- Collins, John, 2000, "Preemptive Prevention," *Journal of Philosophy*, 97: 223–34.
- Collins, John, Ned Hall, e L. A. Paul, 2004, "Counterfactuals and Causation: History, Problems, and Prospects," em J. Collins, N. Hall, e L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 1–58.
- Davidson, Donald, 1980a, "Actions, Reasons, and Causes," origem, 1963, em *Essays on Actions and Events*, Oxford: Clarendon Press, pp. 3–19.
- , 1980b, "Causal Relations," origem, 1967, em *Essays on Actions and Events*, Oxford: Clarendon Press, pp. 149–62.
- , 1980c, "The Individuation of Events," origem, 1969, em *Essays on Actions and Events*, Oxford: Clarendon Press, pp. 163–80.
- , 1980d, "Mental Events," origem, 1970, em *Essays on Actions and Events*, Oxford: Clarendon Press pp. 207–27.
- , 1985, "Reply to Quine on Events," em E. LePore e B. McLaughlin (eds.), *Actions and Events: Essays on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell, pp. 172–6.
- , 1993, "Thinking Causes," em J. Heil e A. Mele (eds.), *Mental Causation*, Oxford: Clarendon Press, pp. 3–17.
- Dowe, Phil, 1992, "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science*, 59: 195–216.
- , 2000, *Physical Causation*, Cambridge: Cambridge University Press.
- , 2001, "A Counterfactual Theory of Prevention and "Causation" by Omission," *Australasian Journal of Philosophy*, 79: 216–26.

Dretske, Fred, 1977, "Referring to Events," em P. French, T. Uehling, Jr., e H. Wettstein (eds.), *Midwest Studies in Philosophy II*, Minneapolis: University of Minnesota Press, pp. 90–9.

Ducasse, C. J., 1926, "On the Nature and Observability of the Causal Relation," *Journal of Philosophy*, 23: 57–68.

Dummett, Michael, 1964, "Bringing About the Past," *Philosophical Review*, 73: 338–59. Eells, Ellery, 1991, *Probabilistic Causality*, Cambridge: Cambridge University Press.

Ehring, Douglas, 1986, "The Transference Theory of Causality," *Synthese*, 67: 249–58.

—, 1997, *Causation and Persistence*, Oxford: Oxford University Press.

Ellis, Brian, 1999, "Causal Powers and Laws of Nature," em H. Sankey (ed.), *Causation and Laws of Nature*, Dordrecht: Kluwer Academic Publishers, pp. 19–34.

Fair, David, 1979, "Causation and the Flow of Energy," *Erkenntnis*, 14: 219–50.

Fales, Evan, 1990, *Causation and Universals*, London: Routledge Press.

Ganeri, Jonardon, Paul Noordhof, and Murali Ramachandran, 1996, "Counterfactuals and Preemptive Causation," *Analysis*, 56: 219–25.

Gasking, Douglas, 1955, "Causation and Recipes," *Mind*, 64: 479–86.

Gödel, Kurt, 1949, "A Remark about the Relationship between Relativity Theory and Idealistic Philosophy," em P. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*, La Salle: Open Court, pp. 557–62.

Good, I. J., 1961, "A Causal Calculus I," *British Journal for the Philosophy of Science*, 11: 305–18.

—, 1962, "A Causal Calculus II," *British Journal for the Philosophy of Science*, 12: 43–51. Hall, Ned, 2000, "Causation and the Price of Transitivity," *Journal of Philosophy*, 97: 198–222.

—, 2004, "Two Concepts of Causation" em J. Collins, N. Hall, e L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 181–204.

—, 2007, "Structural Equations and Causation," *Philosophical Studies*, 132: 109–136.

Halpern, Joseph, 2008, "Defaults and Normality in Causal Structures" em G. Brewka e J. Lang (eds.), *Principles of Knowledge Representation and Reasoning*, Palo Alto: AAAI Press, pp. 198–208.

- Halpern, Joseph e Judea Pearl, 2005, "Causes and Explanations: A Structural-Model Approach— Part 1: Causes," *British Journal for the Philosophy of Science*, 56: 843–87.
- Hart, H. L. A. e A. M. Honore, 1985, *Causation in the Law*, orig. 1959. Oxford: Clarendon Press. Hausman, Daniel, 1998, *Causal Asymmetries*, Cambridge: Cambridge University Press.
- Hitchcock, Christopher, 1993, "A Generalized Probabilistic Theory of Causal Relevance," *Synthese*, 97: 335–64.
- , 1995a, "The Mishap at Reichenbach Fall: Singular vs. General Causation," *Philosophical Studies*, 78: 257–91.
- , 1995b, "Salmon on Explanatory Relevance," *Philosophy of Science*, 62: 304–320.
- , 1996, "The Role of Contrast in Causal and Explanatory Claims," *Synthese*, 107: 395–419.
- , 2001, "The Intransitivity of Causation Revealed in Equations and Graphs," *Journal of Philosophy*, 98: 273–99.
- , 2004, "Do All and Only Causes Raise the Probabilities of Effects?" em J. Collins, N. Hall, e L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 403–18.
- , 2007a, "Prevention, Preemption, and the Principle of Sufficient Reason," *Philosophical Review*, 116: 495–532.
- , 2007b, "What"s Wrong with Neuron Diagrams?" em J. K. Campbell, M. O'Rourke, e H. S. Silverstein (eds.), *Causation and Explanation*, Cambridge, MA: The MIT Press, pp. 69–92.
- , 2011, "Trumping and Contrastive Causation," *Synthese*, 181: 227–40. Horwich, Paul, 1987, *Asymmetries in Time*, Cambridge, MA: The MIT Press.
- Hume, David, 1975, *An Enquiry Concerning Human Understanding*, origem. 1748. Oxford: Clarendon Press. Kant, Immanuel, 1965, *Critique of Pure Reason*, origem, 1781, trans. N. Kemp Smith. New York: Macmillan Press.
- Kim, Jaegwon, 1973, "Causation, Nomic Subsumption, and the Concept of Event," *Journal of Philosophy*, 70: 217–36.
- , 1976, "Events as Property Exemplifications," em M. Brand e D. Walton (eds.), *Action Theory*, Dordrecht: D. Reidel Publishing, pp. 159–77.
- Kistler, Max, 1998, "Reducing Causality to Transmission," *Erkenntnis*, 48: 1–24.

- Kutach, Douglas, 2007, "The Physical Foundations of Causation," em H. Price e R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford: Oxford University Press, pp. 327–50.
- Kvart, Igal, 1986, *A Theory of Counterfactuals*, Indianapolis: Hackett Publishing.
- , 1997, "Cause and Some Positive Causal Impact," em J. Tomberlin (ed.), *Philosophical Perspectives 11: Mind, Causation, and World*, Oxford: Basil Blackwell, pp. 401–32.
- , 2004, "Probabilistic Cause, Edge Conditions, Late Preemption and Discrete Cases," em P. Dowe e P. Noordhof (eds.), *Cause and Chance: Causation in an Indeterministic World*, London: Routledge.
- Lewis, David, 1979, "Counterfactual Dependence and Time's Arrow," *Noûs*, 13: 455–76.
- , 1986a, "Causation," em *Philosophical Papers 2*, Oxford: Oxford University Press, pp. 159–213.
- , 1986b, "Events," em *Philosophical Papers 2*, Oxford: Oxford University Press, pp. 241–69.
- , 2000, "Causation as Influence," *Journal of Philosophy*, 97: 182–97.
- , 2004, "Void and Object," em J. Collins, N. Hall, e L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 277–90.
- Loewer, Barry, 2007, "Counterfactuals and the Second Law," em H. Price e R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford: Oxford University Press, pp. 293–326.
- Mackie, J. L., 1965, "Causes and Conditions," *American Philosophical Quarterly*, 2: 245–64.
- , 1974, *The Cement of the Universe*, Oxford: Oxford University Press.
- Maslen, Cei, 2004, "Causes, Contrasts, and the Nontransitivity of Causation," em J. Collins, N. Hall, e L. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 341–58.
- Maudlin, Tim, 2007, "On the Passing of Time," in *The Metaphysics within Physics*, Oxford: Oxford University Press, pp. 104–42.
- McDermott, Michael, 1995, "Redundant Causation," *British Journal for the Philosophy of Science*, 46: 423–44.
- , 2002, "Causation, Influence, and Sufficiency," *Journal of Philosophy*, 99: 84–101.

- McGrath, Sarah, 2005, "Causation by Omission: A Dilemma," *Philosophical Studies*, 123: 125–48. Mellor, D. H., 1981, *Real Time*, Cambridge: Cambridge University Press.
- , 1988, "On Raising the Chances of Effects," em J. Fetzer (ed.), *Probability and Causality*, Dordrecht: D. Reidel Publishing, pp. 229–40.
- , 1995, *The Facts of Causation*, London: Routledge Press.
- Menzies, Peter, 1989a, "A Unified Account of Causal Relations," *Australasian Journal of Philosophy*, 67: 59–83.
- , 1989b, "Probabilistic Causation and Causal Processes: A Critique of Lewis," *Philosophy of Science*, 56: 642–63.
- , 1996, "Probabilistic Causation and the Pre-emption Problem," *Mind*, 105: 85–117.
- , 2004, "Difference-Making in Context," em J. Collins, N. Hall, em L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press, pp. 139–80.
- , 2007, "Causation in Context," em H. Price e R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford: Oxford University Press, pp. 191–223.
- Menzies, Peter e Huw Price, 1993, "Causation as a Secondary Quality," *British Journal for the Philosophy of Science*, 44: 187–203.
- Mill, J. S., 1846, *A System of Logic*, New York: Harper & Brothers.
- Moore, Michael, 2009, *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*, Oxford: Oxford University Press.
- Noordhof, Paul, 1999, "Probabilistic Causation, Preemption and Counterfactuals," *Mind*, 108: 95–125. Northcott, Robert, 2008, "Causation and Contrast Classes," *Philosophical Studies*, 139: 111–23.
- Papineau, David, 1993, "Can We Reduce Causal Direction to Probabilities?," em Hull, D. M. Forbes, e K. Okruhlik (eds.), *PSA 1992 (Volume 2)*, East Lansing: Philosophy of Science Association, pp. 238–52.
- Paul, L. A., 1999, "Keeping Track of the Time: Emending the Counterfactual Analysis of Causation," *Analysis*, 58: 191–8.
- , 2000, "Aspect Causation," *Journal of Philosophy*, 97: 223–34.
- Paul, L. A. e Ned Hall, 2013, *Causation: A User's Guide*, Oxford: Oxford University Press. Pearl, Judea, 2000, *Causality*, Cambridge: Cambridge University Press.

- Price, Huw, 1991, "Agency and Probabilistic Causality," *British Journal for the Philosophy of Science*, 42: 157–76.
- , 1996, *Time's Arrow and Archimedes' Point*, Oxford: Oxford University Press.
- , 2007, "Causal Perspectivalism," em H. Price e R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford: Oxford University Press, pp. 250–92.
- Quine, W. V. O., 1966, *The Ways of Paradox*, New York: Random House.
- , 1985, "Events and Reification," em E. LePore e B. McLaughlin (eds.), *Actions and Events: Essays on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell, pp. 162–71.
- Ramachandran, Murali, 1997, "A Counterfactual Analysis of Causation," *Mind*, 106: 604–13.
- Russell, Bertrand, 1992, "On the Notion of Cause," origem, 1912, em J. Slater (ed.), *The Collected Papers of Bertrand Russell v6: Logical and Philosophical Papers 1909–1913*, London: Routledge Press, pp. 193–210.
- , 1948, *Human Knowledge: Its Scope and Limits*, New York: Simon and Schuster.
- Salmon, Wesley, 1984, *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.
- , 1997, "Causality and Explanation: A Reply to Two Critiques," *Philosophy of Science*, 64: 461–77.
- , 1998, *Causality and Explanation*, Oxford: Oxford University Press.
- Sartorio, Carolina, 2005, "Causes as Difference-Makers," *Philosophical Studies*, 123: 71–96.
- Sayre, Kenneth, 1977, "Statistical Models of Causal Relations," *Philosophy of Science*, 44: 203–14.
- Schaffer, Jonathan, 2000a, "Trumping Preemption," *Journal of Philosophy*, 97: 165–81.
- , 2000b, "Overlappings: Probability-Raising without Causation," *Australasian Journal of Philosophy*, 78: 40–6.
- , 2000c, "Causation by Disconnection," *Philosophy of Science*, 67: 285–300.
- , 2001, "Causes as Probability-Raisers of Processes," *Journal of Philosophy*, 98: 75–92.
- , 2005, "Contrastive Causation," *Philosophical Review*, 114: 327–58.

—, 2007, “Review: Dowe and Noordhof’s *Cause and Chance*,” *The British Journal for the Philosophy of Science*, 58: 869–74.

—, 2012, “Causal Contextualisms,” em M. Blaauw (ed.), *Contrastivism in Philosophy*, London: Routledge, pp. 35–63.

Shoemaker, Sydney, 1980, “Causality and Properties,” em P. van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*, Dordrecht, D. Reidel Publishing, pp. 109–35.

—, 1998, “Causal and Metaphysical Necessity,” *Pacific Philosophical Quarterly*, 79: 59–77.

Skyrms, Brian, 1984, “EPR: Lessons for Metaphysics,” em P. French, T. Uehling, Jr., e H. Wettstein (eds.), *Midwest Studies in Philosophy IX*, Minneapolis: University of Minnesota Press, pp. 245–55.

Sober, Elliott, 1985, “Two Concepts of Cause,” em P. Asquith e P. Kitcher (eds.), *PSA 1984 (Volume 2)*, East Lansing: Philosophy of Science Association, pp. 405–24.

Spirtes, Peter, Clark Glymour, e Richard Scheines, 1993, *Causation, Prediction, and Search*, New York: Springer-Verlaag.

Strawson, P. F., 1985, “Causality and Explanation,” em B. Vermazen e M. Hintikka (eds.), *Essays on Davidson: Actions and Events*, Oxford: Clarendon Press, pp. 115–36.

Steward, Helen, 1997, *The Ontology of Mind*, Oxford: Clarendon Press.

Suppes, Patrick, 1970, *A Probabilistic Theory of Causality*, Amsterdam: North Holland Publishing. Swain, Marshall, 1978, “A Counterfactual Analysis of Event Causation,” *Philosophical Studies*, 34: 1–19.

Taylor, Richard, 1966, *Action and Purpose*, Upper Saddle River: Prentice Hall.

Tooley, Michael, 1987, *Causation: A Realist Approach*, Oxford: Clarendon Press.

—, 2004, “Probability and causation,” em P. Dowe e P. Noordhof (eds.), *Cause and Chance: Causation in an Indeterministic World*, London: Routledge, pp. 77–119.

Van Fraassen, Bas, 1980, *The Scientific Image*, Oxford: Oxford University Press.

Vendler, Zeno, 1984, “Agency and Causation,” em *Midwest Studies in Philosophy*, 9: 371–84.

Von Wright, G. H., 1975, “On the Logic and Epistemology of the Causal Relation,” em *Causation and Conditionals*, Oxford: Oxford University Press, pp. 95–113.

Weslake, Brad, forthcoming, "A Partial Theory of Actual Causation," *The British Journal for the Philosophy of Science*,

Woodward, James, 1984, "A Theory of Singular Causal Explanation," *Erkenntnis*, 21: 231–62. *A Theory of Explanation*, 2003, *A Theory of Explanation*, Oxford: Oxford University Press.

Yablo, Stephen, 2002, "De Facto Dependence," *Journal of Philosophy*, 99: 130–48.

(IV) Consciência*

Autor: Robert Van Gulick

Tradução: Marco Aurélio Alves & Marcelo Fischborn

Revisão: André Abath

Talvez nenhum aspecto da mente seja mais familiar ou mais enigmático que a consciência e a experiência consciente que temos de nós mesmos e do mundo. O problema da consciência é provavelmente a questão central na teorização atual sobre a mente. Apesar de não haver qualquer acordo nas teorias da consciência, há um consenso difundido, se não universal, de que uma abordagem adequada da mente exige uma compreensão clara da consciência e do seu lugar na natureza. Precisamos entender não apenas o que a consciência é, mas também como se relaciona com outros aspectos, não conscientes, da realidade.

1. História da Questão

É provável que questões sobre a natureza da consciência [*conscious*

* Tradução do verbete “Consciousness” de Robert Van Gulick publicado pela *The Stanford Encyclopedia of Philosophy* (Edição do Verão de 2011), edição de Edward N. Zalta, URL = <http://plato.stanford.edu/archives/sum2011/entries/consciousness/>. Publicado com a autorização do editor Prof. Dr. Edward Zalta. Tradução primeiramente publicada em *Investigação Filosófica*: vol. E2, artigo digital 2, 2012. Disponível em: <<https://periodicos.unifap.br/index.php/investigacaofilosofica/article/view/4855/2182>>.

The following is the translation of the entry on Consciousness by Robert Van Gulick, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <http://plato.stanford.edu/archives/sum2011/entries/consciousness/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at a <http://plato.stanford.edu/entries/consciousness/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and publish this entry.

awareness] sejam levantadas desde tanto tempo quanto têm existido os seres humanos. As práticas de sepultamento neolíticas parecem expressar crenças espirituais, e dão indícios primitivos de pensamento ao menos minimamente reflexivo sobre a natureza da consciência humana (Pearson 1999, Clark and Riel-Salvatore 2001). De modo similar, tem-se descoberto que culturas não-letradas adotam invariavelmente alguma forma de concepção espiritual, ou ao menos animista, que indica certo grau de reflexão sobre a natureza da consciência.

Não obstante, alguns defenderam que a consciência, tal como a conhecemos hoje, é um desenvolvimento histórico relativamente recente, que surgiu em algum momento posterior ao período homérico (Jeynes 1974). De acordo com essa concepção, os seres humanos anteriores, incluindo aqueles que lutaram na Guerra de Tróia, não experienciaram a si mesmos como sujeitos internos unificados de seus próprios pensamentos e ações, ao menos não do modo como o fazemos hoje. Outros alegaram que, mesmo durante o período clássico, não havia nenhuma palavra do grego antigo que correspondesse a "consciência" (Wilkes 1984, 1988, 1995). Embora os antigos tivessem muito a dizer sobre questões da mente, é menos claro que tivessem quaisquer conceitos ou preocupações sobre o que entendemos hoje por consciência.

Embora as palavras "consciente" [*conscious*] e "consciência" [*conscience*] sejam usadas de modo bem diferente hoje em dia, é provável que a ênfase da Reforma na última, como uma fonte interna de verdade, tenha desempenhado algum papel na interiorização tão característica da concepção moderna do eu. O Hamlet que andava pelos palcos em 1600 já via o seu mundo e a si mesmo com olhos profundamente modernos.

Aproximadamente no início da era moderna, no século XVII, a consciência alcançou plenamente o centro do pensamento sobre a mente. De fato, da metade do século XVII ao final do século XIX, a consciência foi amplamente tratada como essencial ou definidora do mental. René Descartes definiu a própria noção de pensamento (*pensée*) em termos de consciência reflexiva ou autoconsciência. Nos *Princípios da Filosofia* (1640), ele escreveu ""pela palavra "pensamento" ("*pensée*"), entendo tudo aquilo de que estamos conscientes como operando em nós.""

Mais tarde, perto do final do século XVII, John Locke fazia uma afirmação similar, embora um pouco mais qualificada, no *Ensaio sobre o Entendimento Humano* (1688):

Não digo que não há alma em um homem porque ele não está ciente [*sensible*] dela durante o sono. Mas digo que em nenhum momento, acordado ou dormindo, ele pode pensar sem estar ciente disso. Estar ciente não é necessário para nada, exceto nossos pensamentos, e para eles isso é, e sempre será, necessário.

Locke recusava-se explicitamente a fazer qualquer hipótese sobre a base substancial da consciência e sua relação com a matéria, mas tratou-a claramente como essencial ao pensamento, bem como à identidade pessoal.

G. W. Leibniz, contemporâneo de Locke, possivelmente inspirado por seu trabalho matemático sobre diferenciação e integração, ofereceu uma teoria da mente no *Discurso de Metafísica* (1686), que permitia infinitos graus de consciência, e talvez mesmo de pensamentos inconscientes, as chamadas “pequenas percepções” [*petite perceptions*]. Leibniz foi o primeiro a distinguir explicitamente entre percepção e apercepção, isto é, grosseiramente, entre consciência e autoconsciência. Na *Monadologia* (1720), ele também ofereceu sua famosa analogia do moinho, para expressar sua crença de que a consciência não poderia surgir da mera matéria. Ele convidou seu leitor a imaginar alguém caminhando pelo interior de um cérebro expandido, tal como caminharia no interior de um moinho, observando todas as suas operações mecânicas, o que para Leibniz esgotava sua natureza física. Em nenhum lugar, afirma ele, esse observador iria ver quaisquer pensamentos conscientes.

Apesar do reconhecimento de Leibniz da possibilidade de pensamento inconsciente, na maior parte dos dois séculos seguintes os domínios do pensamento e da consciência foram tratados como mais ou menos o mesmo. A psicologia associacionista, seja a buscada por Locke, ou depois no século XVIII por David Hume (1739), ou no século XIX por James Mill (1829), visava descobrir os princípios pelos quais os pensamentos ou ideias conscientes interagem ou afetavam um ao outro. John Stuart Mill, filho de James Mill, continuou o trabalho de seu pai na psicologia associacionista, mas permitiu que combinações de ideias pudessem produzir resultantes que estivessem para além de suas partes mentais constituintes, formulando assim um modelo primitivo da emergência do mental (1865).

A abordagem puramente associacionista foi criticada no fim do século XVIII por Immanuel Kant (1787), que argumentou que uma teoria adequada da experiência e da consciência fenomenal exigia uma estrutura muito mais rica da

organização mental e intencional. A consciência fenomenal, segundo Kant, não poderia ser uma mera sucessão de ideias associadas, mas teria de ser no mínimo a experiência de um eu consciente situado em um mundo objetivo estruturado com respeito a espaço, tempo e causalidade.

No mundo anglo-americano as abordagens associacionistas continuaram a ser bastante influentes no século XX, tanto na filosofia como na psicologia, enquanto que na esfera germânica e europeia houve um maior interesse na estrutura mais ampla da experiência, que em parte levou ao estudo da fenomenologia, através do trabalho de Edmund Husserl (1913, 1929), Martin Heidegger (1927), Maurice Merleau-Ponty (1945) e outros que expandiram o estudo da consciência aos reinos social, corporal e interpessoal.

No começo da psicologia científica moderna, na metade do século XIX, a mente ainda era amplamente igualada à consciência e os métodos introspectivos dominavam o campo, como no trabalho de Wilhelm Wundt (1897), Hermann von Helmholtz (1897), William James (1890) e Alfred Titchener (1901). Entretanto, a relação entre consciência e cérebro permanecia um grande mistério, como expresso na observação famosa de T. H. Huxley:

Como algo tão notável quanto um estado de consciência resulta da irritação de tecidos nervosos é simplesmente tão inexplicável quanto o surgimento do gênio, quando Aladim esfregou sua lâmpada (1866).

O início do século vinte viu o eclipse da consciência pela psicologia científica, especialmente nos Estados Unidos, com o surgimento do behaviorismo (Watson 1924, Skinner 1953), embora movimentos como a psicologia Gestalt a tenham mantido em curso como uma questão de interesse científico na Europa (Köhler 1929, Köffka 1935). Nos anos 1960, o domínio do behaviorismo enfraquecia com o surgimento da psicologia cognitiva e sua ênfase no processamento de informação e na criação de modelos de processos mentais internos (Neisser 1965, Gardiner 1985). Entretanto, apesar da ênfase renovada na explicação de capacidades cognitivas como memória, percepção e compreensão da linguagem, a consciência permaneceu como um tópico enormemente negligenciado por mais algumas décadas.

Nos anos 1980 e 90 houve um grande ressurgimento da pesquisa científica e filosófica sobre a natureza e a base da consciência (Baars 1988, Dennett 1991, Penrose 1989, 1994, Crick 1994, Lycan 1987, 1996, Chalmers 1996). Uma vez que a consciência estava de novo em discussão, houve uma

rápida proliferação de pesquisas, com uma enxurrada de livros e artigos, e também o início de revistas especializadas (*The Journal of Consciousness Studies*, *Consciousness and Cognition*, *Psyche*), sociedades profissionais (Association for the Scientific Study of Consciousness - ASSC) e conferências anuais dedicadas exclusivamente à sua investigação (Toward a Science of Consciousness, ASSC).

2. Conceitos da Consciência

As palavras "consciente" e "consciência" são termos genéricos que abrangem uma ampla variedade de fenômenos mentais. Ambas são usadas com uma diversidade de significados, e o adjetivo "consciente" é heterogêneo em sua extensão, sendo aplicado tanto a organismos inteiros – consciência da criatura – como a estados e processos mentais particulares – consciência do estado (Rosenthal 1986, Gennaro 1995, Carruthers 2000).

2.1. A Consciência da Criatura

Um animal, uma pessoa ou outro sistema cognitivo pode ser tratado como consciente em vários sentidos diferentes.

Senciência [*sentience*]. Pode ser consciente no sentido genérico de ser simplesmente uma criatura *senciente*, uma criatura capaz de sentir e responder a seu mundo (Armstrong 1981). Ser consciente nesse sentido pode admitir graus, e exatamente qual tipo de capacidades sensoriais é suficiente pode não estar nitidamente definido. Peixes são conscientes no aspecto relevante? E quanto a camarões e abelhas?

Vigília. Pode-se, ademais, exigir que o organismo esteja efetivamente exercendo essas capacidades, e não meramente que tenha a habilidade ou disposição de exercê-las. Assim, poderíamos considerá-lo consciente apenas se estivesse *acordado e em estado normal de alerta*. Nesse sentido, organismos não seriam considerados conscientes quando adormecidos ou em quaisquer dos níveis mais profundos de coma. Novamente, os limites podem ser nebulosos, e casos intermediários podem estar envolvidos. Por exemplo, está-se consciente no sentido relevante quando se está sonhando, hipnotizado ou em estado de fuga [*fugue state*]?

Autoconsciência. Um terceiro e ainda mais exigente sentido poderia definir as criaturas conscientes como aquelas que estão não apenas conscientes, mas também conscientes de que estão conscientes, tratando assim

a consciência como uma forma de *autoconsciência* (Carruthers 2000). A exigência de autoconsciência poderia ser interpretada de maneiras variadas, e de acordo com isso variará quais criaturas se qualificariam como conscientes no sentido relevante. Se tomamo-la como envolvendo autoconsciência conceitual explícita, muitos animais não-humanos e até crianças jovens poderiam não se qualificar como tal, mas se apenas formas implícitas mais rudimentares de autoconsciência são exigidas, então uma ampla variedade de criaturas não linguísticas poderiam ser consideradas autoconscientes.

Como é [what it is like]. O famoso critério “*como é*” de Thomas Nagel (1974) visa capturar uma outra, e talvez mais subjetiva, noção de ser um organismo consciente. De acordo com Nagel, um ser é consciente apenas se há “algo como é” ser essa criatura, isto é, algum modo subjetivo com o qual o mundo pareça ou em que se apresente do ponto de vista mental ou experiencial da criatura. No exemplo de Nagel, morcegos são conscientes porque há algo que é como ser um morcego que experiencia seu mundo através de seus sentidos ecolocalizadores, apesar de nós humanos, do nosso ponto de vista humano, não podermos entender empaticamente como tal modo de consciência é do ponto de vista do próprio morcego.

Sujeito de estados conscientes. Uma quinta alternativa seria definir a noção de um organismo consciente em termos de estados conscientes. Isto é, poderíamos primeiro definir o que torna um estado mental um estado mental consciente, e então definir uma criatura como sendo consciente em termos da posse desses estados. Nosso conceito de um organismo consciente dependeria, assim, da abordagem específica que damos aos estados conscientes (seção 2.2).

Consciência transitiva. Além de descrever criaturas como conscientes nesses vários sentidos, há também sentidos relacionados em que criaturas são descritas como sendo *conscientes de várias coisas*. Essa distinção é por vezes marcada como uma distinção entre noções *transitivas* e *intransitivas* de consciência, com a primeira envolvendo algum objeto ao qual a consciência é dirigida (Rosenthal 1986).

2.2. A Consciência de Estados

A noção de um estado mental consciente também tem uma variedade de sentidos distintos, embora talvez inter-relacionados. Há ao menos seis opções principais.

Estados em que se sabe que se está [states one is aware of]. Em uma leitura comum, um estado mental consciente é simplesmente um estado mental que se sabe estar nele (Rosenthal 1986, 1996). Estados conscientes, nesse sentido, envolvem uma forma de *meta-mentalidade* ou *meta-intencionalidade*, na medida em que exigem estados mentais que são eles mesmos a respeito de estados mentais. Ter um desejo consciente de uma xícara de café é ter esse desejo e também estar simultaneamente e diretamente consciente de que se tem esse desejo. Pensamentos e desejos inconscientes, nesse sentido, são simplesmente aqueles que temos sem saber [*without being aware*] que temos, quer nossa falta de autoconhecimento resulte de simples falta de atenção ou de causas psicanalíticas mais profundas.

Estados qualitativos. Estados podem também ser tratados como conscientes em um sentido aparentemente muito diferente e mais *qualitativo*. Isto é, poderíamos considerar consciente um estado somente se tem ou envolve propriedades qualitativas ou experienciais do tipo ao qual frequentemente nos referimos como “qualia” ou “qualidades sensoriais primárias” [*raw sensory feels*]. (Ver o verbete por Tye 2009.) A percepção que alguém tem do Merlot que está bebendo ou do tecido que está examinando conta como um estado mental consciente nesse sentido, porque envolve vários qualia sensoriais, por exemplo, qualia de sabor no caso do vinho e qualia de cores na experiência visual que se tem do pano. Há um considerável desacordo sobre a natureza de tais qualia (Churchland 1985, Shoemaker 1990, Clark 1993, Chalmers 1996) e igualmente sobre sua existência. Tradicionalmente, os qualia foram tratados como características monádicas da experiência, privadas e inefáveis, mas as teorias atuais dos qualia frequentemente rejeitam ao menos algum desses comprometimentos (Dennett 1990).

Estados fenomenais. Por vezes refere-se a esses qualia como propriedades fenomenais e ao tipo associado de consciência como *consciência fenomenal*, mas o último termo talvez seja mais propriamente aplicado à estrutura completa da experiência e envolva muito mais que os qualia sensoriais. A estrutura fenomenal da consciência também abrange muito da organização espacial, temporal e conceitual de nossa experiência do mundo e de nós mesmos como agentes nele (ver seção 4.3). Por isso, é provavelmente melhor, ao menos inicialmente, distinguir o conceito de consciência fenomenal do de consciência qualitativa, embora sem dúvida eles se sobreponham.

Estados como-é [*what-it-is-like*]. A consciência em ambos esses sentidos liga-se também à noção de Thomas Nagel (1974) de uma criatura

consciente, na medida em que se considera um estado mental consciente no sentido de “*como é*” simplesmente se há algo que seja como estar nesse estado. O critério de Nagel poderia ser entendido como buscando fornecer uma concepção em primeira pessoa, ou interna, do que torna um estado um estado fenomenal ou qualitativo.

Consciência de acesso. Os estados poderiam ser conscientes em um sentido de acesso, aparentemente muito diferente, que tem mais a ver com as relações intramentais. A este respeito, um estado ser consciente é uma questão de sua disponibilidade para interagir com outros estados e do acesso que se tem ao seu conteúdo. Nesse sentido mais funcional, que corresponde ao que Ned Block (1995) chama de *consciência de acesso*, um estado visual ser consciente não é tanto uma questão de ter ou não um “*como é*” [*what it's likeness*] qualitativo, mas de ele e a informação visual que carrega estarem ou não geralmente disponíveis para uso e orientação do organismo. Na medida em que a informação nesse estado está rica e flexivelmente disponível para o organismo que a contém, então ele conta como um estado consciente no aspecto relevante, quer ou não tenha qualquer percepção [*feel*] qualitativa ou fenomenal no sentido de Nagel.

Consciência narrativa. Os estados também poderiam ser considerados conscientes em um sentido *narrativo* que apela à noção de um “fluxo de consciência”, considerado como uma narração de episódios progressiva, mais ou menos serial, da perspectiva de um eu efetivo ou meramente virtual. A ideia seria igualar os estados mentais conscientes da pessoa com aqueles que aparecem no fluxo (Dennett 1991, 1992).

Embora essas seis noções sobre o que torna consciente um estado possam ser especificadas independentemente, elas obviamente não o podem sem conexões potenciais, e nem esgotam o reino das opções possíveis. Extraíndo conexões, poderíamos defender que estados aparecem no fluxo de consciência apenas na medida em que estamos conscientes [*aware*] deles, e assim forjar um vínculo entre a primeira noção meta-mental de um estado consciente e o conceito de um fluxo ou narrativa. Ou poderíamos conectar a noção de acesso com a noção fenomenal ou qualitativa de um estado consciente, tentando mostrar que estados que representam daquele modo tornam seus conteúdos amplamente disponíveis no aspecto exigido pela noção do acesso.

Buscando ir além das seis opções, poderíamos distinguir estados conscientes de nãoconscientes apelando a aspectos de sua dinâmica

intramental e de suas interações outras que não as meras relações de acesso; por exemplo, estados conscientes poderiam manifestar um estoque mais rico de interações sensíveis ao conteúdo, ou um grau maior de orientação intencional flexível, do tipo associado com o controle autoconsciente do pensamento. De modo alternativo, poderíamos tentar definir estados conscientes em termos de criaturas conscientes. Isto é, poderíamos dar uma explicação [*account*] do que é ser uma criatura consciente ou talvez até mesmo de um eu consciente, e então definir a noção de um estado consciente em termos de ser um estado dessa criatura ou sistema, o que seria o inverso da última opção considerada acima para definir criaturas conscientes em termos de estados mentais conscientes.

2.3. Consciência como uma Entidade

O substantivo “consciência” tem uma variedade igualmente diversificada de significados, que em grande medida são paralelos àqueles do adjetivo “consciente”. Podem ser feitas distinções entre a consciência de criaturas e a de estados, bem como entre as variedades de cada uma. Pode-se referir especificamente à consciência fenomenal, consciência de acesso, consciência reflexiva ou meta-mental e à consciência narrativa, entre outras variedades.

Aqui a consciência em si não é tratada tipicamente como uma entidade substancial, mas meramente como a reificação abstrata de qualquer propriedade ou aspecto que é atribuído pelo uso relevante do adjetivo “consciente”. Consciência de acesso é simplesmente a propriedade de ter o tipo exigido de relações internas de acesso, e a consciência qualitativa é simplesmente a propriedade que é atribuída quando “consciente” é aplicado, no sentido qualitativo, a estados mentais. O quanto isso por si só nos compromete com o estatuto ontológico da consciência dependerá de quão platônico se é a respeito de universais em geral (ver o verbete por Klima 2008). Isso não precisa nos comprometer com a consciência como uma entidade distinta mais do que o uso de “quadrado”, “vermelho” ou “suave” nos compromete com a existência da quadradeza [*squareness*], vermelhidão ou suavidade como entidades distintas.

Embora não seja uma norma, poderíamos ainda assim adotar uma concepção realista mais robusta da consciência como um componente da realidade. Isto é, poderíamos pensar a consciência como algo mais parecido com campos eletromagnéticos do que com a vida.

Desde a morte do vitalismo, não pensamos a vida *por si só* como algo distinto de coisas vivas. Há coisas vivas, incluindo organismos, estados,

propriedades e partes de organismos, comunidades e linhagens evolutivas de organismos, mas a vida não é ela mesma mais uma coisa, um componente adicional da realidade, alguma força vital que se acrescenta às coisas vivas. Aplicamos corretamente os adjetivos “vivo” e “animado” [*alive*] a muitas coisas, e ao fazer isso poderíamos dizer que lhes estamos atribuindo vida, mas sem nenhum significado ou realidade além daquele envolvido em serem coisas vivas.

Campos eletromagnéticos, por contraste, são tratados como partes reais e independentes do nosso mundo físico. Mesmo que por vezes se possa especificar os valores de um certo campo apelando ao comportamento das partículas em seu interior, os campos são tratados eles mesmos como constituintes concretos da realidade, e não meramente como abstrações ou conjuntos de relações entre partículas.

De modo similar, poderíamos tratar “consciência” como se referindo a um componente ou aspecto da realidade que manifesta a si mesmo em estados e criaturas conscientes, mas que é mais do que meramente a nominalização abstrata do adjetivo “consciente” que lhes aplicamos. Embora essas concepções fortemente realistas não sejam muito comuns atualmente, elas devem ser incluídas no espaço lógico de opções.

Há assim muitos conceitos de consciência, e tanto “consciente” como “consciência” são usados de modos muito variados, sem um significado privilegiado ou canônico. No entanto, isso pode ser menos um problema do que uma excessiva fartura. A consciência é uma característica complexa do mundo, e entendê-la exigirá uma diversidade de ferramentas conceituais para lidar com seus vários aspectos diferentes. A pluralidade conceitual é, assim, justamente o que se esperaria. Na medida em que se evita a confusão, sendo claro quanto a seus significados, há um grande valor em se ter uma variedade de conceitos pelos quais podemos acessar e apreender a consciência em toda a sua rica complexidade. No entanto, não se deve supor que a pluralidade conceitual acarreta divergência de referência. Nossos múltiplos conceitos de consciência podem, de fato, apanhar aspectos variados de um mesmo e unificado fenômeno mental subjacente. Se assim o é, e em que medida, permanece uma questão em aberto.

3. Problema da Consciência

A tarefa de entender a consciência é um projeto igualmente diversificado. Não só muitos aspectos diferentes da mente contam como

conscientes em algum sentido, mas cada um é também aberto a vários aspectos em que poderia ser explicado ou modelado. Entender a consciência envolve não apenas uma multiplicidade de coisas a serem explicadas [*explananda*], mas também de questões que elas colocam e tipos de respostas que exigem. Sob risco simplificar excessivamente, as questões relevantes podem ser reunidas sob três rubricas cruas como as questões de o que é, como é e de por que existe:

- A questão descritiva: *O que é a consciência?* Quais são suas características principais? E por que meios elas podem ser mais bem descobertas, descritas e modeladas?
- A questão explicativa: *Como a consciência do tipo relevante passa a existir?* Ela é um aspecto primitivo da realidade e, se não, como a consciência no aspecto relevante surge (ou poderia surgir), ou é causada, por entidades ou processos não-conscientes?
- A questão funcional: *Por que a consciência do tipo relevante existe?* Ela tem uma função e, se sim, qual? Ela age causalmente e, se sim, quais seus tipos de efeitos? Ela faz diferença para a operação dos sistemas em que está presente e, se sim, por que e como?

As três questões focam respectivamente em descrever as características da consciência, explicar sua base ou causa subjacente e explicar sua função ou valor. A divisão em três é certamente um tanto artificial e, na prática, as respostas que se dá a cada uma dependerão em parte do que se diz sobre as outras. Não se pode, por exemplo, responder adequadamente à questão do que é a consciência e descrever suas características principais sem abordar a questão do porquê, de seu papel funcional nos sistemas cujas operações afeta. Nem se poderia explicar como o tipo relevante de consciência poderia surgir de processos não-conscientes a menos que se tivesse uma abordagem clara de precisamente quais características teriam de ser causadas ou realizadas para contar como a produzindo. Apesar dessas ressalvas, a divisão em três vias de questões fornece uma estrutura útil para articular o projeto explicativo global, e para avaliar a adequação de teorias ou modelos particulares da consciência.

4. A Questão Descritiva: *Quais são as Características da Consciência?*

A questão do *que é a consciência* nos pede para descrever e modelar as características principais da consciência, mas exatamente quais características são relevantes variará de acordo com o tipo de consciência que

pretendermos capturar. As propriedades principais da consciência de acesso podem ser completamente diferentes daquelas da consciência qualitativa ou fenomenal, e as da consciência reflexiva ou narrativa podem diferir de ambas. Entretanto, acumulando teorias detalhadas de cada tipo, podemos ter a esperança de encontrar ligações importantes entre elas, e talvez até descobrir que coincidem em ao menos alguns aspectos-chave.

4.1. Dados de Primeira e de Terceira Pessoa

O projeto descritivo geral exigirá uma variedade de métodos de investigação (Flanagan 1992). Embora pudéssemos tratar ingenuamente os fatos da consciência como bastante autoevidentes para exigir quaisquer métodos sistemáticos de obtenção de dados, a tarefa epistêmica está realmente longe de ser trivial (Husserl 1913).

O acesso introspectivo de primeira pessoa proporciona uma fonte de conhecimento [*insight*] de nossa vida mental consciente, mas não é por si só suficiente nem mesmo útil, a menos que usado de modo treinado e disciplinado. Obter os indícios necessários da estrutura da experiência exige tanto que nos tornemos observadores fenomenologicamente sofisticados quanto que complementemos nossos resultados introspectivos com vários tipos de dados de terceira pessoa disponíveis para o observador externo (Searle 1992, Varela 1995, Siewert 1998).

Como os fenomenólogos sabem há mais de um século, descobrir a estrutura da experiência consciente demanda um rigoroso olhar introspectivo [*inner directed stance*] que é completamente diferente de nossa forma cotidiana de autoconsciência (Husserl 1929, Merleau-Ponty 1945). A observação habilidosa do tipo necessário exige treino, esforço e a capacidade de adotar perspectivas alternativas sobre a própria experiência.

A necessidade de dados empíricos de terceira pessoa obtidos por observadores externos é talvez mais óbvia com respeito aos tipos claramente mais funcionais de consciência, como a consciência de acesso, mas é exigida mesmo com respeito à consciência fenomenal e qualitativa. Por exemplo, estudos de déficit que correlacionam várias regiões de danos neurais e funcionais com anomalias da experiência consciente podem tornar-nos cientes de aspectos da estrutura fenomenal que escapam à nossa consciência introspectiva normal. Como esses estudos de caso mostram, coisas que parecem inseparavelmente unidas ou singulares, do nosso ponto de vista normal

de primeira pessoa, podem tornar-se separadas na experiência (Sacks 1985, Shallice 1988, Farah 1995).

Ou, para apanhar um outro exemplo, dados de terceira pessoa podem nos tornar conscientes de como nossas experiências de estar agindo e nossas experiências da duração de eventos afetam umas as outras, de um modo que nunca poderíamos discernir através da mera introspecção (Libert 1985, Wegner 2002). E esses fatos obtidos por métodos de terceira pessoa nem são meramente a respeito das causas ou bases da consciência; eles frequentemente dizem respeito à estrutura mesma da própria consciência fenomenal. Métodos interativos de primeira pessoa, terceira pessoa e talvez até de segunda pessoa (Varela 1995) serão todos exigidos para coletar a evidência necessária.

Usando todas essas fontes de dados, estaremos, espera-se, aptos a construir modelos descritivos detalhados dos vários tipos de consciência. Embora características específicas da maior importância possam variar entre os diferentes tipos, nosso projeto descritivo global precisará enfrentar ao menos os sete aspectos gerais da consciência a seguir (seções 4.2-4.8).

4.2. O Caráter Qualitativo

O *caráter qualitativo* é frequentemente igualado às chamadas “sensações brutas” [*raw feels*], e ilustrado pela vermelhidão que se experimenta quando se olha para tomates maduros ou o sabor doce específico que se encontra ao provar um abacaxi igualmente maduro (Locke 1968). O tipo relevante de caráter qualitativo não está restrito a estados sensoriais, mas é tipicamente tomado como estando presente enquanto um aspecto dos estados experienciais em geral, tais como os pensamentos ou desejos experienciados (Siewert 1998).

Para alguns, a existência dessas sensações parece demarcar o ponto a partir do qual estados ou criaturas são realmente conscientes. Se um organismo sente e responde de maneira adequada a seu mundo, mas carece de tais qualia, então no máximo poderia contar como consciente em um sentido frouxo e não-literal. Ou ao menos assim pareceria àqueles que tomam a consciência qualitativa, no sentido de “como é”, como sendo central filosófica e cientificamente (Nagel 1974, Chalmers 1996).

Problemas de variadas formas sobre os qualia se avultaram no passado recente – Pode haver qualia invertidos? (Block 1980a, 1980b, Shoemaker 1981, 1982) Os qualia são epifenomenais? (Jackson 1982, Chalmers 1996) Como os

estados neurais podem originar qualia? (Levine 1983, McGinn 1991) Mas a questão do que é a consciência levanta um problema mais básico sobre os qualia, a saber, o de dar uma descrição clara e articulada de nosso espaço de qualia e do estatuto de seus qualia específicos.

Na ausência de tal modelo, erros factuais ou descritivos são muito prováveis. Por exemplo, afirmações sobre a ininteligibilidade da ligação entre o vermelho experienciado e qualquer substrato neural possível dessa experiência tratam, às vezes, o quale da cor relevante como uma propriedade simples e *sui generis* (Levine 1983), mas a vermelhidão fenomenal existe de fato em um espaço complexo de cores, com múltiplas dimensões sistemáticas e relações de semelhança (Hardin 1992). Entender o quale específico da cor em relação a essa ampla estrutura relacional não nos dá apenas uma melhor apreensão descritiva da sua natureza qualitativa, mas pode também fornecer-nos alguns “ganchos”, aos quais poderíamos anexar as ligações psicofísicas inteligíveis.

A cor pode ser a exceção quanto a termos um entendimento formal específico e bem desenvolvido do espaço qualitativo relevante, mas provavelmente não é uma exceção com respeito à importância desses espaços para o nosso entendimento das propriedades qualitativas em geral (Clark 1993, P. M. Churchland 1995). (Ver o verbete “qualia”, por Tye 2009)

4.3. A Estrutura Fenomenal

A *estrutura fenomenal* não deve ser confundida com a estrutura qualitativa, apesar do uso por vezes intercambiável de “qualia” e “propriedades fenomenais” na literatura. A “organização fenomenal” abrange todos os vários tipos de ordem e estrutura encontrados no domínio da experiência, isto é, no domínio do mundo tal como nos *aparece*. Obviamente há ligações importantes entre o fenomenal e o qualitativo. Com efeito, os qualia talvez fossem mais bem entendidos como propriedades dos objetos fenomenais ou experienciados, mas de fato há bem mais para o fenomenal do que sensações brutas [*raw feels*]. Como Kant (1787), Husserl (1913) e gerações de fenomenólogos mostraram, a estrutura fenomenal da experiência é ricamente intencional e envolve não apenas ideias e qualidades sensoriais, mas representações complexas de tempo, espaço, causa, do corpo, de si mesmo, do mundo e da estrutura organizada da realidade vivida, em todas as suas formas conceituais e nãoconceituais.

Uma vez que muitos estados não-conscientes também têm aspectos

intencionais e representacionais, pode ser melhor considerar a estrutura fenomenal como envolvendo um tipo especial de organização e conteúdo intencional e representacional, o tipo distintivamente associado à consciência (Siewert 1998). (Ver o verbete sobre “teorias representacionais da consciência”, por Lycan 2008).

Responder a questão do que é a consciência exige uma abordagem cuidadosa da estrutura representacional coerente e densamente organizada em que as experiências particulares estão inseridas. Dado que a maior parte dessa estrutura está apenas implícita na organização da experiência, ela não pode simplesmente ser lida por introspecção. Articular a estrutura do domínio fenomenal de modo claro e inteligível é um longo e difícil processo de inferência e construção de modelos (Husserl 1929). A introspecção pode ajudar, mas muita construção teórica e engenhosidade são também necessárias.

4.4. Subjetividade

A *subjetividade* é mais uma noção por vezes igualada na literatura aos aspectos qualitativos ou fenomenais da consciência, mas novamente há boas razões para reconhecê-la, ao menos em algumas de suas formas, como um aspecto distinto da consciência – relacionada ao qualitativo e ao fenomenal, mas diferente de cada um. Em particular, a forma epistêmica da subjetividade diz respeito aos aparentes limites da cognoscibilidade ou mesmo da compreensibilidade de vários fatos sobre a experiência consciente (Nagel 1974, Van Gulick 1985, Lycan 1996).

Na abordagem de Thomas Nagel (1974), os fatos sobre como é ser um morcego são subjetivos no sentido relevante, porque podem ser completamente entendidos apenas a partir do tipo de ponto de vista do morcego. Apenas criaturas capazes de ter ou experienciar de modo similar tais experiências podem entender seu como-é no sentido empático necessário. Fatos sobre a experiência consciente podem ser entendidos no máximo incompletamente de um ponto de vista exterior, de terceira pessoa, tal como os que estão associados com a ciência física objetiva. Uma concepção similar sobre os limites da teoria em terceira pessoa parece estar por traz de afirmações a respeito do que a Maria hipotética de Frank Jackson (1982), a super cientista de cores, não poderia entender sobre experienciar o vermelho, devido à sua própria história empobrecida de experiência visual acromática.

É um debate aberto se os fatos sobre a experiência são realmente

limitados epistemicamente dessa maneira (Lycan 1996), mas a alegação de que entender a consciência exige formas especiais de conhecimento e de acesso a partir do ponto de vista interior é intuitivamente plausível e tem uma longa história (Locke 1688). Assim, qualquer resposta adequada à questão do que é a consciência tem de cobrir o estatuto epistêmico da consciência, ambas as nossas habilidades e limites para compreendê-la (Papineau 2002, Chalmers 2003).

4.5. Organização Autoperspectivista

A estrutura perspectivista da consciência é um aspecto de sua organização fenomenal global, mas é importante o suficiente para merecer discussão por si só. Na medida em que a perspectiva chave é aquela do eu consciente, a característica específica poderia ser chamada de *autoperspectividade*. Experiências conscientes não existem como átomos mentais isolados, mas como modos ou estados de um eu ou sujeito consciente (Descartes 1644, Searle 1992, em contraste com Hume 1739). Uma experiência visual de uma esfera azul é sempre uma questão de haver algum eu ou sujeito para o qual pareceu desse modo. Uma dor forte ou aguda é sempre uma dor sentida ou experienciada por algum sujeito consciente. O eu não precisa aparecer como um elemento explícito em nossas experiências, mas, como Kant (1787) notou, o “eu penso” tem de acompanhar ao menos potencialmente cada uma delas.

O eu poderia ser tomado como o ponto de vista perspectivista a partir do qual o mundo de objetos está presente para a experiência (Wittgenstein 1921). Ele fornece não apenas uma perspectiva espacial e temporal para nossa experiência do mundo, mas também de significado e inteligibilidade. A coerência intencional do domínio experiencial repousa sobre a interdependência dual entre o eu e o mundo: o eu como perspectiva a partir da qual os objetos são conhecidos e o mundo como a estrutura integrada de objetos e eventos cujas possibilidades de serem experienciados implicitamente definem a natureza e a localização do eu (Kant 1787, Husserl 1929).

Os organismos conscientes obviamente diferem na medida em que constituem um eu unificado e coerente, e provavelmente diferem quanto ao tipo ou grau de foco perspectivista que incorporam em suas respectivas formas de experiência (Lorenz 1977). A consciência pode não exigir um eu distinto ou substancial do tipo tradicional cartesiano, mas ao menos algum grau de

organização perspectivista semelhante a um ego parece essencial para a existência de algo que poderia contar como uma experiência consciente. Parece que as experiências não seriam capazes de existir sem um eu ou sujeito para experienciá-las mais do que poderiam as ondas do oceano sem o mar através do qual se movem. A questão descritiva exige, assim, alguma abordagem do aspecto autoperspectivista da experiência e da organização de tipo egoica [*self-like*] das mentes conscientes da qual depende, mesmo que a abordagem relevante trate o eu de um modo relativamente deflacionário e virtual (Dennett 1991, 1992).

4.6. Unidade

A *unidade* está intimamente ligada à autoperspectiva, mas merece sua própria menção como um aspecto chave da organização da consciência. Sistemas conscientes e estados mentais conscientes envolvem ambos muitas formas diversificadas de unidade. Algumas são unidades causais, associadas à integração da ação e controle em um foco unificado de agência. Outras são formas mais representacionais e intencionais de unidade, envolvendo a integração de diversos itens do conteúdo em ligações de muitas escalas e níveis (Cleeremans 2003).

Algumas dessas integrações são relativamente locais, como quando diversas características detectadas com uma única modalidade sensorial são combinadas na representação de objetos externos portando tais características, por exemplo, quando se tem a experiência visual consciente de uma lata vermelha de sopa passando sobre um guardanapo verde listrado (Triesman and Gelade 1980).

Outras formas de unidade intencional envolvem uma gama muito mais ampla de conteúdos. O conteúdo da experiência atual que alguém tem da sala em que está sentado depende, em parte, de sua localização em uma estrutura bem maior, associada com a consciência de sua existência como observador temporalmente estendido em um mundo de objetos espacialmente conectados e existentes de modo independente (Kant 1987, Husserl 1913). A experiência individual pode ter o conteúdo que tem apenas porque reside nessa estrutura unificada mais ampla de representação. (Ver o verbete “Unidade da consciência”, por Brook e Raymond 2010.)

4.7. Intencionalidade e Transparência

Estados mentais conscientes são tipicamente tratados como tendo um aspecto representacional ou intencional, na medida em que são sobre coisas, referem-se a coisas ou têm condições de satisfação. A experiência visual de alguém *representa corretamente* o mundo se há violetas em um vaso branco sobre a mesa (em oposição a Travis 2004), sua memória consciente é *do* ataque ao World Trade Center e seu desejo consciente é *por* um copo de água gelada. Entretanto, estados não-conscientes também podem exibir intencionalidade desse modo, e é importante entender de que maneiras os aspectos representacionais de estados conscientes se assemelham e diferem dos não-conscientes (Carruthers 2000). Searle (1990) oferece um ponto de vista contrário, de acordo com o qual apenas estados conscientes e disposições para ter estados conscientes podem ser genuinamente intencionais, mas a maioria dos teóricos tratam a intencionalidade como estendendo-se amplamente no interior do domínio inconsciente. (Ver o verbete “Consciência e intencionalidade”, por Siewert 2011).

Uma dimensão potencialmente importante da distinção diz respeito à assim chamada *transparência*, que é uma característica importante da consciência em dois sentidos metafóricos inter-relacionados, cada um dos quais com aspectos intencionais, experienciais e funcionais.

Diz-se frequentemente que a experiência perceptual consciente é transparente ou, na expressão de G. E. Moore (1922), “diáfana”. “Olhamos através” de nossa experiência sensorial transparentemente, na medida em que parecemos diretamente conscientes dos objetos e eventos externos que presenciamos, em vez de estarmos conscientes de quaisquer propriedades da experiência pelas quais ela nos apresenta ou representa tais objetos. Quando olho o vento a soprar no pasto, é da grama verde ondulante que estou consciente, e não de qualquer propriedade verde da minha experiência visual. (Ver o verbete “Teorias representacionais da consciência”, por Lycan 2008.) O próprio Moore acreditava que poderíamos nos tornar conscientes dessas qualidades posteriores com esforço e redirecionamento da atenção, embora alguns defensores contemporâneos da transparência o neguem (Harman 1990, Tye 1995).

Pensamentos e experiências conscientes também são transparentes em um sentido semântico, em que seus significados nos parecem imediatamente conhecidos no próprio ato de pensá-los (Van Gulick 1992). Nesse sentido, poderíamos dizer que “através deles pensamos diretamente” no que significam ou representam. Nesse sentido semântico, a transparência pode corresponder,

ao menos em parte, ao que John Searle chama de “intencionalidade intrínseca” da consciência (Searle 1992).

Nossos estados mentais conscientes parecem ter seus significados intrinsecamente, ou a partir do interior, apenas pelo que são por si mesmos, em contraste com muitas teorias externalistas do conteúdo mental, que baseiam o significado em relações causais, contrafactuais ou informacionais entre seus objetos semânticos ou referenciais.

A concepção de conteúdos conscientes como determinados intrinsecamente e internamente autoevidentes é por vezes apoiada por apelos a intuições sobre o cérebro na cuba, que fazem parecer que os estados mentais conscientes do cérebro encubado manteriam todos os seus conteúdos intencionais normais, apesar da falta de todas as suas ligações causais e informacionais normais com o mundo (Horgan e Tienson 2002). Há permanente controvérsia sobre tais casos e sobre as concepções concorrentes internalistas (Searle 1992) e externalistas (Dretske 1995) da intencionalidade consciente.

Embora a transparência semântica e a intencionalidade intrínseca tenham algumas afinidades, não devem ser simplesmente igualadas, já que é possível acomodar a primeira noção em uma abordagem mais externalista do conteúdo e do significado. Obviamente, tanto a transparência semântica como a sensorial não apenas dizem respeito aos aspetos representacionais e intencionais da consciência, mas também são aspectos experienciais da nossa vida consciente. Ambas também têm aspectos funcionais, na medida em que as experiências conscientes interagem umas com as outras de maneiras ricamente apropriadas ao conteúdo, que manifestam nosso entendimento transparente de seus conteúdos.

4.8. Fluxo Dinâmico

A *dinâmica* da consciência é evidente na ordem coerente de seu processo constante de fluxo e autotransformação, o que William James (1890) chamou de “*fluxo da consciência*”. Algumas sequências temporais da experiência são geradas por fatores puramente internos, como quando se raciocina sobre um quebra-cabeça, e outras dependem em parte de causas externas, como quando se persegue uma bola no ar, mas mesmo as últimas são em grande medida moldadas pelo modo como a consciência transforma a si mesma.

Seja parcialmente em resposta a influências de fora, ou inteiramente do interior, de momento a momento cada sequência da experiência cresce

coerentemente para além daquelas que a precederam, constrangida e permitida pela estrutura global de ligações e limites incorporados em sua organização anterior subjacente (Husserl 1913). A esse respeito, a consciência é um sistema autopoietico, isto é, um sistema autocriador e auto-organizador (Varela e Maturana 1980).

Como um agente mental consciente, posso fazer muitas coisas, tais como vasculhar meu quarto, vasculhar uma imagem mental dele, rever na memória os pratos de uma refeição recente em um restaurante, juntamente com seus sabores e aromas, raciocinar passo a passo sobre um problema complexo, ou planejar uma ida às compras e executar o plano quando chegar ao mercado. Todos esses são hábitos e atividades comuns, mas cada um envolve a geração dirigida de experiências, de uma maneira que manifesta uma compreensão prática implícita de suas propriedades intencionais e conteúdos interconectados (Van Gulick 2000).

A consciência é um processo dinâmico, e assim uma resposta descritiva adequada à questão do que ela é tem de lidar com mais do que apenas suas propriedades estáticas ou momentâneas. Em particular, tem de dar alguma explicação da dinâmica temporal da consciência e da maneira pela qual seu fluxo autotransformante reflete tanto sua coerência intencional quanto a autocompreensão semântica incorporada nos controles organizados através dos quais as mentes conscientes refazem continuamente a si mesmas, enquanto sistemas autopoieticos engajados em seus mundos.

Uma abordagem descritiva exaustiva da consciência precisaria lidar com mais do que apenas essas sete características, mas ter uma abordagem clara de cada uma delas seria andar um longo caminho em direção a responder a questão “O que é a consciência?”.

5. A Questão Explicativa: Como a Consciência Pode Existir?

A questão de *como* a consciência pode existir foca na explicação, antes que na descrição. Pede-nos para explicar o estatuto básico da consciência e seu lugar na natureza. Ela é por si só uma característica fundamental da realidade, ou sua existência depende de outros itens, não-conscientes, sejam eles físicos, biológicos, neurais ou computacionais? Se a segunda opção é o caso, podemos explicar ou entender como os itens não-conscientes relevantes puderam causar ou realizar a consciência? Ou, colocado de modo simples, podemos explicar como fazer algo consciente a partir de coisas que não são conscientes?

5.1. A Diversidade de Projetos Explicativos

A questão de como pode existir a consciência não é uma questão única, mas antes uma família geral de questões mais específicas (Van Gulick 1995). Todas dizem respeito à possibilidade de explicar algum gênero ou aspecto da consciência, mas variam no que explicar [*explananda*], na explicação [*explanans*] e nos critérios para uma explicação bem sucedida. Por exemplo, poderíamos perguntar se podemos explicar a consciência de acesso computacionalmente, simulando as relações de acesso necessárias em um modelo computacional. Ou, em vez disso, poderíamos nos interessar em se as propriedades fenomenais e qualitativas da mente de uma criatura consciente podem ser deduzidas *a priori* de uma descrição das propriedades neurais de seus processos cerebrais. Ambas são versões da questão de como a consciência pode existir, mas perguntam sob perspectivas de projetos explicativos muito diferentes, e assim podem diferir em suas respostas (Lycan 1996). Seria impraticável, se não impossível, catalogar todas as versões possíveis da questão de como a consciência existe, mas algumas das principais opções podem ser listadas.

O que explicar [explananda]. Os possíveis explicáveis incluiriam os vários tipos de consciência de estados e de criaturas distinguidos acima, bem como as sete características da consciência listadas em resposta à questão de o que é a consciência. Esses dois tipos de coisas explicáveis se sobrepõem e interagem. Poderíamos, por exemplo, buscar explicar o aspecto dinâmico tanto da consciência fenomenal como da de acesso. Ou poderíamos tentar explicar a subjetividade tanto da consciência qualitativa como da metamental. Nem toda característica se aplica a cada gênero de consciência, mas todas se aplicam a vários. A maneira de se explicar uma dada característica em relação a um gênero de consciência pode não corresponder ao que é necessário para explicá-la em relação a um outro.

Explicações [explanans]. O espectro de explicações possíveis também é diversificado. Talvez em sua forma mais ampla, a questão de como pode existir a consciência pergunta como a consciência do tipo relevante poderia ser causada ou realizada por itens não-conscientes, mas podemos gerar uma profusão de questões mais específicas, restringindo ainda mais o escopo das explicações relevantes. Poderíamos buscar explicar como uma dada característica da consciência é causada ou realizada por processos *neurais*, estruturas *biológicas*, mecanismos *físicos*, relações *funcionais* ou *telefuncionais*, por uma organização *computacional* ou mesmo por estados *mentais não-conscientes* subjacentes. As

perspectivas para o sucesso explicativo irão variar de acordo. Em geral, quanto mais limitado e elementar o escopo da explicação [*explanans*], mais difícil é o problema de explicar como isso poderia ser suficiente para produzir a consciência (Van Gulick 1995).

Critérios de explicação. O terceiro parâmetro chave é como se define o critério para uma explicação bem sucedida. Poderíamos exigir que o que é explicado seja dedutível *a priori* da explicação, embora seja controverso se isso é um critério necessário ou mesmo suficiente para explicar a consciência (Jackson 1993). Sua suficiência dependerá em parte da natureza das premissas a partir das quais a dedução será feita. Como questão de lógica, serão necessários alguns princípios de ponte para conectar proposições ou frases sobre a consciência com aqueles que não a mencionam. Se as premissas dizem respeito a fatos físicos ou neurais, então serão necessários alguns princípios de ponte ou ligações que conectem esses fatos com fatos sobre a consciência (Kim 1998). Ligações brutas, sejam elas nômicas ou meramente correlações bem confirmadas, poderiam fornecer uma ponte logicamente suficiente para inferir conclusões sobre a consciência. Mas provavelmente não nos permitiriam ver como ou por que tais conexões se dão, e assim estariam aquém de explicar completamente como a consciência existe (Levine 1983, 1993, McGinn 1991).

Poderíamos legitimamente pedir por mais, em particular por alguma abordagem que tornasse inteligível por que essas ligações se dão, e talvez por que não poderiam deixar de ser assim. Um modelo em dois estágios familiar para explicar as macropropriedades em termos de microsubstratos é invocado frequentemente. No primeiro passo, analisa-se a macropropriedade em termos de condições funcionais, e então, no segundo estágio, mostra-se que as microestruturas obedecendo as leis de seu próprio nível é nomicamente suficiente para garantir a satisfação das condições funcionais relevantes (Armstrong 1968, Lewis 1972).

As micropropriedades de coleções de moléculas de H₂O a 20°C são suficientes para satisfazer as condições para a liquidez da água que compõem. Além do mais, o modelo torna inteligível como a liquidez é produzida pelas micropropriedades. Poderia parecer que uma explicação satisfatória de como a consciência é produzida exigiria uma história similar em dois estágios. Sem isso, mesmo a dedutibilidade *a priori* poderia parecer explicativamente menos do que suficiente, embora a necessidade de uma tal história permaneça uma questão controversa (Block e Stalnaker 1999, Chalmers e Jackson 2001).

5.2. A Lacuna Explicativa

Seguindo Joseph Levine (1983), nossa incapacidade atual de suprir uma ligação inteligível adequada é por vezes descrita como a existência de uma *lacuna explicativa*, e como indicando nosso entendimento incompleto de como a consciência poderia depender de um substrato não-consciente, especialmente um substrato físico. A afirmação básica de que existe uma lacuna admite muitas variações de generalidade e, portanto, de força.

Talvez em sua forma mais fraca, ela asseve um limite *prático* em nossas capacidades explicativas *atuais*; dadas as nossas teorias e modelos atuais, não podemos articular agora uma ligação inteligível. Uma versão mais forte faz uma afirmação *em princípio* sobre as nossas *capacidades humanas*, asserindo assim que, dados os nossos limites cognitivos humanos, nunca seremos capazes de ultrapassar a lacuna. Para nós, ou para criaturas cognitivamente semelhantes a nós, deve permanecer um mistério residual (McGinn 1991). Colin McGinn (1995) defendeu que, dada a natureza inerentemente espacial, tanto de nossos conceitos perceptuais quanto dos conceitos científicos que deles derivamos, nós, humanos, não somos conceitualmente ajustados para entender a natureza da ligação psicofísica. Fatos sobre essa ligação estão tão cognitivamente fechados a nós quanto os fatos sobre multiplicação ou sobre raízes quadradas estão para os tatus. Eles não se enquadram em nosso repertório conceitual e cognitivo. Uma versão ainda mais forte da asserção da lacuna remove a restrição à nossa natureza cognitiva e nega *em princípio* que a lacuna possa ser preenchida por *qualquer agente cognitivo*.

Aqueles que defendem a existência da lacuna discordam entre si sobre que conclusões metafísicas, se alguma, seguem-se dos nossos supostos limites epistêmicos. O próprio Levine relutou em extrair quaisquer conclusões ontológicas antifisicalistas (Levine 1993, 2001). Por outro lado, alguns neodualistas têm tentado usar a existência da lacuna para refutar o fisicalismo (Foster 1996, Chalmers 1996). Quanto mais forte a premissa epistêmica, maior a esperança de se derivar uma conclusão metafísica. Assim, não surpreendentemente, conclusões dualistas são frequentemente sustentadas por apelos à suposta impossibilidade *em princípio* de se fechar a lacuna.

Se pudéssemos ver, em bases *a priori*, que não há como a consciência ser inteligivelmente explicada como surgindo do físico, então não seria um grande passo concluir que isso não pode ser o caso (Chalmers 1996). Entretanto, a própria força dessa alegação epistêmica torna difícil chegar, sem incorrer em petição de princípio, ao resultado metafísico. Assim, aqueles que

desejam usar a alegação forte da existência de uma lacuna *em princípio* para refutar o fisicalismo têm de encontrar bases independentes para sustentá-la. Alguns apelaram para argumentos de conceitabilidade [*conceivability*] em suporte, tal como a presumida conceitabilidade de zumbis molecularmente idênticos a humanos conscientes, mas desprovidos de toda consciência fenomenal (Campbell 1970, Kirk 1974, Chalmers 1996). Outros argumentos em suporte invocam a suposta natureza não-funcional da consciência, e daí sua alegada resistência ao método científico padrão de explicar propriedades complexas (por exemplo, dominância genética) em termos de condições funcionais realizadas fisicamente (Block 1980a, Chalmers 1996). Esses argumentos antifisicalistas evitam incorrer em petição de princípio, mas repousam, eles próprios, em alegações e intuições que são controversas e não completamente independentes do ponto de vista básico que se tem sobre o fisicalismo. Discussões sobre o tópico continuam ativas e em andamento.

Nossa incapacidade atual de ver qualquer maneira de preencher a lacuna pode exercer alguma atração em nossas intuições, mas pode simplesmente refletir os limites de nossa teorização atual, antes que uma barreira intransponível em princípio (Dennett 1991). Além do mais, alguns fisicalistas defenderam que lacunas explicativas devem ser esperadas, e são até implicadas, por versões plausíveis do fisicalismo ontológico, versões essas que tratam os agentes humanos como sistemas cognitivos fisicamente realizados, com limites inerentes que derivam de sua origem evolutiva e do seu modo contextual e situado de entendimento (Van Gulick 1985, 2003; McGinn 1991, Papineau 1995, 2002). Nessa perspectiva, antes de refutar o fisicalismo, a existência de lacunas explicativas pode confirmá-lo. A discussão e o desacordo nesses tópicos permanecem ativos e em andamento.

5.3. Explicação Reducionista e não-Reducionista

Como a necessidade de uma ligação inteligível já mostrou, uma dedutibilidade *a priori* não é, por si só, claramente suficiente para uma explicação bem sucedida (Kim 1980), nem é claramente necessária. Alguma ligação lógica mais fraca poderia ser suficiente em muitos contextos explicativos. Por vezes, contando o suficiente de uma história sobre como os fatos de um tipo dependem dos de outro podemos ficar satisfeitos de que os últimos de fato causam ou realizam os primeiros, mesmo que não possamos estritamente deduzir completamente os primeiros dos últimos.

A dedução interteórica estrita foi tomada como norma redutiva pela abordagem empirista lógica da unidade da ciência (Putnam e Oppenheim 1958), mas nas décadas mais recentes uma imagem não-reducionista mais frouxa das relações entre as várias ciências tem ganhado preferência. Em particular, materialistas não-reducionistas têm defendido a assim chamada “autonomia das ciências especiais” (Fodor 1974) e a ideia de que entender o mundo natural exige que usemos uma diversidade de sistemas conceituais e representacionais, que podem não ser estritamente intertraduzíveis ou capazes de serem postos na correspondência estrita exigida pelo antigo paradigma dedutivo das relações entre níveis (Putnam 1975).

A economia é frequentemente citada como um exemplo (Fodor 1974, Searle 1992). Fatos econômicos podem ser realizados por processos físicos subjacentes, mas ninguém pede seriamente que sejamos capazes de deduzir os fatos econômicos relevantes das descrições detalhadas de suas bases físicas subjacentes, ou que sejamos capazes de colocar os conceitos e o vocabulário da economia em correspondência estrita com aqueles das ciências físicas.

Apesar disso, nossa incapacidade dedutiva não é vista como motivo para suspeitas ontológicas; não há um problema da relação “dinheiro-matéria”. Tudo o que exigimos é algum entendimento geral e menos que dedutivo de como propriedades e relações econômicas poderiam ter por base propriedades e relações físicas. Assim, poderíamos optar por um critério similar para interpretar a questão de como a consciência pode existir e o que conta como explicar como a consciência poderia ser causada ou realizada por itens não-conscientes.

No entanto, alguns críticos, tais como Kim (1987), têm contestado a coerência de qualquer concepção que pretenda ser tanto não-redutiva quanto fiscalista, embora defensores dessas concepções tenham por sua vez replicado (Van Gulick 1993).

Outros têm defendido que a consciência é especialmente resistente à explicação em termos físicos, devido às diferenças inerentes entre nossos modos de compreensão subjetivos e objetivos. Thomas Nagel (1974) ficou famoso ao defender que existem limites inevitáveis impostos à nossa capacidade de entender a fenomenologia da experiência do morcego devido à nossa incapacidade de tomar empaticamente uma perspectiva experiencial como aquela que caracteriza a experiência auditiva ecolocalizadora que o morcego tem de seu mundo. Dada nossa incapacidade de submeter-nos a uma experiência similar, podemos ter no máximo um entendimento parcial da natureza de tal experiência. Nenhuma quantidade de conhecimento colhida pela

perspectiva objetiva externa de terceira pessoa das ciências naturais será supostamente suficiente para nos permitir entender o que o morcego pode entender de sua própria experiência a partir do seu ponto de vista subjetivo interno de primeira pessoa.

5.4. Expectativas de Sucesso Explicativo

A questão de como a consciência pode existir subdivide-se, assim, em uma família diversificada de questões mais específicas, dependendo do tipo ou característica específica da consciência que se busca explicar, das restrições específicas que se faz quanto ao escopo das explicações [*explanans*] e ao critério que se usa para definir o sucesso explicativo. Algumas das variantes resultantes parecem mais fáceis de responder que outras. O progresso pode parecer provável em alguns dos assim chamados “problemas fáceis” da consciência, tais como explicar a dinâmica da consciência de acesso em termos da organização funcional ou computacional do cérebro (Baars 1988). Outros podem parecer menos tratáveis, especialmente o assim chamado “problema difícil” (Chalmers 1995), que é mais ou menos o de dar uma abordagem inteligível que nos permita ver de uma maneira intuitivamente satisfatória como a consciência fenomenal ou “como é” pode surgir de processos físicos ou neurais no cérebro.

Respostas positivas a algumas versões das questões de como a consciência pode existir parecem ao alcance da mão, mas outras parecem permanecer profundamente frustrantes. Nem deveríamos supor que toda versão tem uma resposta positiva. Se o dualismo é verdadeiro, então, ao menos em alguns de seus tipos, a consciência poderia ser básica e fundamental. Se assim o for, não seremos capazes de explicar como ela surge de itens nãoconscientes, dado que ela simplesmente não surge dessa forma.

A concepção que adotamos sobre as expectativas de se explicar a consciência dependerá tipicamente da perspectiva adotada. Fisicalistas otimistas provavelmente verão os lapsos explicativos atuais como mero reflexo de um estágio inicial de investigação, a serem certamente remediados num futuro não tão distante (Dennett 1991, Searle 1992, P. M. Churchland 1995). Para os dualistas, aqueles mesmos impasses significarão a falência do programa fisicalista e a necessidade de reconhecer a consciência como por si só um constituinte fundamental da realidade (Robinson 1982, Foster 1989, 1996, Chalmers 1996). O que se vê depende em parte de onde se está, e o projeto em

andamento de explicar a consciência será acompanhado por contínuo debate sobre o seu estatuto e as expectativas de sucesso.

6. A Questão Funcional: *Por Que* a Consciência Existe?

A questão funcional, ou do *porquê*, pergunta sobre o *valor* ou *papel* da consciência, e assim pergunta indiretamente sobre a sua origem. Ela tem uma *função* e, se tem, qual é? Ela faz alguma diferença na operação dos sistemas onde está presente e, se faz, por que e como faz? Se a consciência existe como uma característica complexa de sistemas biológicos, então seu valor adaptativo é provavelmente relevante na explicação de sua origem evolutiva, ainda que, evidentemente, sua função atual, se tem alguma, não precisa ser a mesma que ela talvez teve quando surgiu. Funções adaptativas frequentemente mudam no decorrer do tempo biológico. Questões sobre o valor da consciência também têm uma dimensão *moral* em ao menos dois sentidos. Somos inclinados a ver o estatuto moral de um organismo como ao menos parcialmente determinado pela natureza e extensão em que é consciente, e estados conscientes, principalmente estados conscientes afetivos, tais como prazeres e dores, desempenham um papel central em muitas das abordagens do valor que subjazem a teoria moral (Singer 1975).

Assim como as questões do que é e como pode existir, a pergunta de por que a consciência existe também coloca um problema geral que se subdivide em uma variedade de indagações mais específicas. Na medida em que as várias formas de consciência, tais como consciência de acesso, fenomenal ou metamental, são distintas e separáveis – o que permanece uma questão aberta – elas provavelmente também diferem em seus papéis e valores específicos. Portanto, a questão de por que a consciência existe pode muito bem não ter uma resposta única ou uniforme.

6.1. Estatuto Causal da Consciência

Talvez a questão mais básica colocada por qualquer versão da pergunta de por que existe é se a consciência do tipo relevante tem ou não algum impacto causal. Se ela não tem nenhum efeito e não faz nenhuma diferença causal, seja ela qual for, então ela pareceria incapaz de desempenhar qualquer papel significativo nos sistemas e organismos onde está presente, eliminando assim, de partida, a maior parte das questões relativas ao seu possível valor. A ameaça de irrelevância epifenomenal também não poderia, dessa forma, ser

simplesmente descartada como uma não-opção óbvia, uma vez que pelo menos algumas formas de consciência são seriamente defendidas na literatura recente como não tendo nenhum caráter causal. (Ver o verbete sobre “epifenomenalismo”, por Robinson 2011.) Esses problemas são levantados principalmente em relação aos qualia e à consciência qualitativa (Huxley 1874, Jackson 1982, Chalmers 1996), mas desafios também são apresentados ao estatuto causal das outras, incluindo a consciência metamental (Velmans 1991).

Argumentos tanto metafísicos quanto empíricos são oferecidos para sustentar ambas as posições. Entre os primeiros, estão aqueles que apelam a intuições sobre a conceitabilidade e a possibilidade lógica de zumbis, ou seja, de seres cujo comportamento, organização funcional e estrutura física até o nível molecular são idênticos a agentes humanos normais, mas que não possuem nenhum qualia ou consciência qualitativa. Alguns (Kirk 1970, Chalmers 1996) afirmam que tais seres são possíveis em mundos que compartilham todas as nossas leis físicas, mas outros negam isso (Dennett 1991, Levine 2001). Se são possíveis em tais mundos, então parece seguir-se que, mesmo no nosso mundo, os qualia não afetam o curso dos eventos físicos, incluindo os que constituem nossos comportamentos humanos. Se esses eventos se desenrolam da mesma forma, estejam ou não presentes os qualia, então os qualia parecem ser inertes ou epifenomenais, ao menos em relação aos eventos do mundo físico. Porém, esses argumentos e as intuições sobre zumbis nas quais se assentam são controversos, e a correção [*soundness*] deles continua em disputa (Searle 1992, Yablo 1998, Balog 1999).

Argumentos de um tipo bem mais empírico têm contestado o estatuto causal da consciência metamental, ao menos na medida em que sua presença pode ser medida pela habilidade de descrever o estado mental de alguém. Argumenta-se que evidências científicas mostram que uma consciência desse tipo não é nem necessária para nenhum tipo de habilidade mental nem ocorre suficientemente cedo para servir de causa dos atos ou processos tipicamente tomados como seus efeitos (Velmans 1991). De acordo com aqueles que defendem esses argumentos, os tipos de habilidades mentais que normalmente são vistas como necessitando a consciência podem todas ser realizadas inconscientemente sem a autoconsciência supostamente necessária.

Ademais, mesmo quando a consciência autorreflexiva está presente, ela supostamente ocorre tarde demais para ser a causa das ações relevantes, ao invés de seu resultado ou, no máximo, um efeito conjunto de alguma causa anterior compartilhada (Libet 1985). A autoconsciência ou consciência

metamental, de acordo com esses argumentos, mostra-se um efeito psicológico posterior ao invés de uma causa iniciadora, mais como uma impressão *post facto* ou o resultado exibido na tela do computador do que as efetivas operações do processador que produziram tanto a resposta do computador quanto o que está na tela.

Mais uma vez, os argumentos são controversos, e tanto os supostos dados quanto sua interpretação estão sujeitos a intensos desacordos (ver Flanagan 1992, e comentários acompanhando Velmans 1991). Apesar de os argumentos empíricos, assim como o dos zumbis, exigirem que consideremos seriamente se algumas formas de consciência são ou não causalmente menos potentes do que normalmente se presume, muitos teóricos veem os dados empíricos como não sendo uma ameaça real ao caráter causal da consciência.

Se os epifenomenalistas estão errados e a consciência, em suas várias formas, é de fato causal, então que tipos de efeitos ela tem e que diferença faz? Como os processos mentais que envolvem o tipo relevante de consciência diferem daqueles que não o envolvem? Quais funções a consciência poderia desempenhar? As próximas seis seções (6.2 – 6.7) discutem algumas das respostas mais comuns. Apesar das várias funções se sobreporem até certo ponto, cada uma delas é distinta, e elas diferem também quanto aos tipos de consciência com os quais cada uma está mais adequadamente ligada.

6.2. Controle Flexível

Flexibilidade e sofisticação de controle aumentadas. Processos mentais conscientes parecem fornecer formas de controle altamente flexíveis e adaptativas. Ainda que processos automáticos inconscientes possam ser extremamente eficientes e rápidos, eles normalmente operam de maneira mais fixa e predeterminada do que aqueles que envolvem autoconsciência (Anderson 1983). A consciência é, portanto, da maior importância quando estamos lidando com situações novas e problemas ou demandas não previamente encontrados (Penfield 1975, Armstrong 1981).

As explicações convencionais da aquisição de habilidades ressaltam a importância da consciência durante a fase inicial de aprendizado, que gradualmente dá lugar a processos mais automáticos, do tipo que requer pouca atenção ou supervisão consciente (Schneider e Shiffrin 1977). O processamento consciente permite a construção ou compilação de rotinas especificamente preparadas a partir de unidades elementares, bem como o controle deliberado

de suas execuções.

Há uma troca bem conhecida entre flexibilidade e velocidade; processos conscientes controlados adquirem sua versatilidade sob medida às custas de serem lentos e cansativos, em contraste com a rapidez fluida das operações mentais inconscientes automáticas (Anderson 1983). Os crescimentos relevantes na flexibilidade parecem mais intimamente conectados com o tipo metamental, ou de ordem superior [*higher-order*], de consciência, na medida em que a melhora da habilidade de controlar processos depende do aumento de autoconsciência. No entanto, flexibilidade e modos sofisticados de controle também podem ser associados às formas de consciência fenomenal e de acesso.

6.3. Coordenação Social

Capacidade aumentada para coordenação social. A consciência de tipo metamental pode envolver não apenas um acréscimo de autoconsciência, mas também uma melhora na compreensão dos estados mentais de outras criaturas com mentes, principalmente de outros membros do grupo social (Humphreys 1982). Criaturas que são conscientes no sentido metamental relevante não apenas possuem crenças, motivações, percepções e intenções, mas também compreendem como é ter tais estados e são conscientes tanto delas mesmas como de outros como os possuindo.

Esse acréscimo de conhecimento mutuamente compartilhado da mente um do outro permite que os organismos relevantes interajam, cooperem e comuniquem-se de formas muito mais avançadas e adaptativas. Apesar de a consciência metamental ser o tipo mais obviamente associado a essa função socialmente coordenativa, a consciência narrativa, do tipo associado ao fluxo de consciência, também é certamente relevante, na medida em que envolve a aplicação na própria criatura de habilidades interpretativas que surgem em parte de suas aplicações sociais (Ryle 1949, Dennett 1978, 1992).

6.4. Representação Integrada

Representação mais unificada e densamente integrada da realidade. A experiência consciente nos apresenta um mundo de objetos existindo independentemente no espaço e no tempo. Esses objetos nos são apresentados tipicamente de uma forma multimodal, que envolve a integração de informações de vários canais sensoriais, assim como do conhecimento de base e da memória.

A experiência consciente não nos apresenta propriedades ou características isoladas, mas sim objetos e eventos situados num mundo independente em fluxo, e ela faz isso incorporando em sua organização e dinâmica experiencial a densa rede de relações e interconexões que coletivamente constituem a estrutura significativa de um mundo de objetos (Kant 1787, Husserl 1913, Campbell 1997).

Evidentemente, nem toda informação sensorial precisa ser experienciada para ter um efeito adaptativo no comportamento. Conexões adaptativas sensorio-motoras não-experienciais podem ser encontradas tanto em organismos simples quanto em alguns dos processos mais diretos e reflexivos de organismos superiores. Mas quando há experiência, ela fornece uma representação mais unificada e integrada da realidade, que normalmente propicia vias mais abertas de resposta (Lorenz 1977). Considere, por exemplo, a representação do espaço em um organismo cujos canais de entrada sensorial são simplesmente ligados ao movimento ou à orientação de poucos mecanismos fixos, tais como aqueles para alimentação ou para agarrar presas, e compare isso com um organismo capaz de usar sua informação espacial para a navegação flexível do seu ambiente e para quaisquer outros objetivos ou metas espacialmente relevantes que possa ter, como quando uma pessoa examina visualmente seu escritório ou sua cozinha (Gallistel 1990).

É a representação desse último tipo que normalmente é disponibilizada pelo modo integrado de apresentação associado à experiência consciente. A unidade do espaço experienciado é apenas um exemplo do tipo de integração associado à nossa consciência de um mundo objetivo. (Ver o verbete sobre a “unidade da consciência”, por Brook e Raymont 2010.)

Esse papel ou valor integrativo é mais obviamente associado com o tipo fenomenal de consciência, mas, na medida em que os qualia desempenham um papel na apresentação de objetos na experiência, esse papel unificador também é associado à consciência nos sentidos qualitativo e de “como-é” [*what it's like*]. Ademais, ele está intimamente ligado à transparência da experiência, principalmente a transparência semântica, descrita em resposta à questão sobre o que é a consciência (Van Gulick 1993).

6.5. Acesso Informacional

Mais acesso informacional global. A informação carregada por estados mentais conscientes normalmente está disponível para o uso de diversos

subsistemas mentais e para a aplicação em um amplo rol de situações e ações potenciais (Baars 1988). A informação não-consciente é mais passível de ser encapsulada no interior de módulos mentais particulares e ficar disponível apenas para o uso concernente às aplicações diretamente conectadas às operações desse subsistema (Fodor 1983). Tornar uma informação consciente normalmente amplia a esfera de sua influência e o número de formas em que pode ser usada para guiar e conformar adaptativamente tanto o comportamento interno quanto externo. O caráter consciente de um estado pode ser em parte uma questão daquilo que Dennett chamou de “celebridade cerebral”, ou seja, a capacidade desse estado de ter um impacto compatível com o conteúdo de outros estados mentais.

Essa função particular está ligada mais diretamente, e por definição, à noção de consciência de acesso (Block 1995), mas a consciência metamental, bem como as formas fenomenal e qualitativa, parecem todas associadas a esse crescimento na disponibilidade de informação (Armstrong 1981, Tye 1985).

6.6. Livre-Arbítrio

Aumento da liberdade de escolha ou livre-arbítrio. A questão do livre-arbítrio permanece um problema filosófico perene, não apenas no que diz respeito à sua existência ou não, mas até mesmo no que se refere a em que isso poderia ou deveria consistir (Dennett 1984, van Inwagen 1983, Hasker 1999, Wegner 2002). (Ver o verbete sobre “Livre-arbítrio”, por O'Connor 2011.) Talvez a própria noção de livre-arbítrio continue muito obscura e controversa para esclarecer o que quer que seja sobre o papel da consciência, mas há uma intuição tradicional de que as duas coisas estão intimamente ligadas.

A consciência é vista como abrindo um reino de possibilidades, uma esfera de opções na qual o eu consciente pode escolher ou agir livremente. A consciência parece, no mínimo, uma pré-condição necessária para qualquer forma de liberdade ou autodeterminação desse tipo (Hasker 1999). Como alguém poderia exercer o tipo exigido de livre-arbítrio e, ao mesmo tempo, permanecer apenas no interior do domínio inconsciente? Como alguém poderia determinar sua própria vontade sem ser consciente dela e das opções disponíveis para moldá-la?

A liberdade para escolher a ação a tomar e a habilidade de determinar a própria natureza e desenvolvimento futuro podem admitir diversos graus e variações interessantes, ao invés de serem simplesmente uma questão de tudo

ou nada, e as várias formas ou níveis de consciência podem estar associadas a graus e tipos correspondentes de liberdade e autodeterminação (Dennett 1984, 2003). A ligação com a liberdade parece mais forte para a forma metamental de consciência, dada a sua ênfase na autoconsciência, mas conexões potenciais também parecem igualmente possíveis para a maior parte das outras formas.

6.7. Motivação Intrínseca

Estados intrinsecamente motivacionais. Ao menos alguns estados conscientes parecem ter intrinsecamente a força motora que têm. Em particular, os papéis funcionais e motivacionais de estados afetivos conscientes, tais como prazeres e dores, parecem intrínsecos ao seu caráter experiencial e inseparáveis de suas propriedades qualitativas e fenomenais, ainda que isso tenha sido colocado em questão (Nelkin 1989, Rosenthal 1991). O aspecto motivacional positivamente atrativo do prazer parece uma parte de sua sensação fenomenal diretamente experienciada, assim como o caráter afetivo negativo da dor, ao menos no caso da experiência não-patológica normal.

Há considerável controvérsia sobre o quanto o sentimento e a força motora da dor podem ser dissociados em casos anormais, e alguns negam a própria existência de tais aspectos intrinsecamente motivacionais (Dennett 1991). No entanto, ao menos no caso normal, a força motivacional negativa da dor *parece* constitutiva da sensação da própria experiência.

Exatamente *como* isso poderia ser o caso continua não sendo muito claro, e talvez a aparente força motivacional intrínseca e diretamente experienciada seja ilusória. Mas se for real, então pode ser um dos sentidos mais importantes e evolutivamente mais antigos em que a consciência faz diferença para os sistemas e processos mentais onde está presente (Humphreys 1992).

Outras sugestões foram feitas sobre as possíveis funções e valor da consciência, e essas seis evidentemente não exaurem as opções. No entanto, elas estão entre as hipóteses recentes mais proeminentes, e fornecem um levantamento satisfatório dos tipos de respostas que têm sido oferecidas à questão de por que a consciência existe por aqueles que acreditam que a consciência de fato faz alguma diferença.

6.8. Papéis Constitutivos e Contingentes

Um ponto adicional exige esclarecimento sobre os vários sentidos em que as funções propostas poderiam responder à pergunta de por que a

consciência existe. Em particular, deve-se distinguir entre casos *constitutivos* e casos de *realização contingente*. No primeiro, desempenhar o papel constitui ser consciente no sentido relevante, enquanto que no segundo a consciência de um dado tipo é apenas uma entre várias formas em que o papel exigido poderia ser realizado (Van Gulick 1993).

Por exemplo: tornar a informação disponível globalmente para uso de uma ampla variedade de subsistemas e aplicações comportamentais pode constituir seu caráter consciente no sentido de acesso. Por outro lado, mesmo que as formas qualitativas e fenomenais de consciência envolvam uma representação altamente unificada e densamente integrada da realidade objetiva, pode ser possível produzir representações que tenham essas características funcionais, mas que não sejam de natureza qualitativa ou fenomenal.

O fato de que *em nós* os modos de representação com essas características também tenham propriedades qualitativas e fenomenais pode refletir fatos históricos contingentes sobre a solução particular de modelo que calhou surgir em nossa linhagem evolutiva. Se isso for verdade, deve haver meios bem diversos de se atingir um resultado comparável sem a consciência qualitativa ou fenomenal. Não é claro que essa seja a forma adequada de se pensar a consciência qualitativa e fenomenal; talvez a associação com representações unificadas e densamente integradas seja de fato tão íntima e constitutiva quanto parece ser o caso na consciência de acesso (Carruthers 2000). Independentemente de como essa questão seja resolvida, é importante não misturar abordagens constitutivas com abordagens de realização contingente quando se investiga a função da consciência e se responde à questão de por que ela existe (Chalmers 1996).

7. Teorias da Consciência

Em resposta às perguntas sobre o que é, como é e por que existe, muitas teorias da consciência foram propostas nos últimos anos. Entretanto, nem todas as teorias da consciência são teorias da mesma coisa. Elas variam não apenas quanto aos tipos de consciência que têm como seu objeto, mas também em suas finalidades teóricas.

Talvez a maior divisão seja entre teorias metafísicas gerais, que pretendem situar a consciência no interior de um esquema ontológico abrangente da realidade, e teorias específicas, que oferecem explicações

detalhadas de sua natureza, características e função. A linha separando esses dois tipos de teoria é um tanto obscura, principalmente na medida em que muitas teorias específicas carregam ao menos alguns comprometimentos com questões metafísicas mais gerais. No entanto, é útil manter a divisão em mente quando se faz um levantamento do leque de ofertas teóricas correntes.

8. Teorias Metafísicas da Consciência

Teorias metafísicas gerais oferecem respostas para a versão consciente do problema mente-corpo: qual o estatuto ontológico da consciência em relação ao mundo da realidade física? Em grande medida, as respostas disponíveis acompanham as alternativas convencionais ao problema mente-corpo, incluindo as principais versões de dualismo e fisicalismo.

8.1. Teorias Dualistas

Teorias dualistas veem ao menos alguns aspectos da consciência como estando fora do reino físico, mas formas específicas de dualismo diferem a respeito de quais são, exatamente, esses aspectos. (Ver o verbete por Robinson 2011.)

O *dualismo de substância*, tal como o dualismo tradicional cartesiano (Descartes 1644), afirma a existência de ambas as substâncias, física e não-física. Essas teorias implicam a existência de mentes ou egos não-físicos como entidades às quais a consciência é inerente. Apesar do dualismo de substância estar em grande medida em desuso no momento, existem alguns defensores contemporâneos (Swinburne 1986, Foster 1989, 1996).

O *dualismo de propriedade*, em suas várias versões, goza de maior popularidade atualmente. Todas as teorias desse tipo afirmam a existência de propriedades conscientes que não são nem idênticas nem redutíveis a propriedades físicas, mas que podem, entretanto, ser instanciadas exatamente pelas mesmas coisas que instanciam propriedades físicas. Nesse sentido, elas podem ser classificadas como teorias de *duplo aspecto*. Elas tomam algumas partes da realidade – organismos, cérebros, estados ou processos neurais – como instanciando propriedades de dois tipos distintos e desconexos: as físicas e as conscientes, fenomenais ou qualitativas. Teorias de duplo aspecto, ou dualistas de propriedades, podem ser de ao menos três tipos diferentes.

O *dualismo de propriedades fundamental* vê propriedades mentais conscientes como constituintes básicos da realidade, ao lado de propriedades

físicas, como carga eletromagnética. Elas podem interagir de maneira causal e legiforme [*law-like*] com outras propriedades fundamentais, tais como as da física, mas ontologicamente a existência delas não é nem dependente nem derivada de nenhuma outra propriedade (Chalmers 1996).

O *dualismo de propriedades emergentes* vê as propriedades conscientes como surgindo de organizações complexas de constituintes físicos, mas de uma forma radical tal que o resultado emergente é algo para além de suas causas físicas, não previsível *a priori* e não explicável em termos de suas naturezas estritamente físicas. A coerência de tais abordagens emergentistas foi desafiada (Kim 1998), mas elas têm defensores (Hasker 1999).

O *dualismo de propriedades monista e neutro* trata tanto propriedades mentais conscientes quanto propriedades físicas como, de alguma forma, dependentes ou derivadas de um nível mais básico de realidade, que não é, ele mesmo, nem mental nem físico (Russell 1927, Strawson 1994). Entretanto, se entendemos o dualismo como a tese de que há dois reinos distintos de entidades ou propriedades fundamentais, então o monismo neutro talvez não devesse ser classificado como uma versão de dualismo de propriedades, uma vez que não vê nem as propriedades mentais nem as físicas como últimas ou fundamentais.

O *pampsiquismo* pode ser visto como um quarto tipo de dualismo de propriedades, uma vez que toma todos os constituintes da realidade como tendo algumas propriedades psíquicas, ou ao menos protopsíquicas, distintas de qualquer propriedade física que possam ter (Nagel 1979). De fato, o monismo neutro poderia ser consistentemente combinado com alguma versão de *pamprotopsiquismo* (Chalmers 1996), na qual os aspectos protomentais dos micro-constituintes poderiam dar origem, em condições de combinação propícias, à consciência em sentido pleno. (Ver o verbete sobre “pampsiquismo”, por Seager e Allen-Hermanson 2012.)

A natureza do aspecto protopsíquico relevante continua obscura e essas teorias enfrentam um dilema, se oferecidas na esperança de responder ao Problema Difícil [*Hard Problem*]. Ou as propriedades protopsíquicas envolvem o tipo de sensação qualitativa fenomenal que gera o Problema Difícil, ou elas não envolvem. Caso envolvam, é difícil entender como elas podem possivelmente ocorrer como propriedades ubíquas da realidade. Como pode um elétron ou um quark ter qualquer sensação experiencial desse tipo? No entanto, se as propriedades protopsíquicas não envolvem sensações desse tipo, então não é claro como seriam, de alguma forma, mais aptas do que as propriedades físicas para dar conta da consciência qualitativa ao solucionar o Problema Difícil.

Muitos argumentos foram apresentados a favor da teoria dualista e de outras teorias antifisicalistas da consciência. Alguns são, em grande parte, *a priori* em natureza, como aqueles que apelam para a suposta conceitabilidade de zumbis (Kirk 1970, Chalmers 1996) ou versões do argumento do conhecimento (Jackson 1982, 1986), que visa alcançar uma conclusão antifisicalista sobre a ontologia da consciência a partir dos aparentes limites em nossa habilidade de compreender inteiramente os aspectos qualitativos da experiência consciente através de abordagens físicas de terceira-pessoa dos processos cerebrais. (Ver Jackson 1998, 2004 para uma visão contrária; ver também os verbetes sobre “zumbis”, por Kirk 2012, e “Qualia: o argumento do conhecimento”, por Nida-Rümelin 2010.) Outros argumentos a favor do dualismo possuem bases mais empíricas, tais como aqueles que apelam à suposta lacuna causal nas cadeias de causação física no cérebro (Eccles e Popper 1977), ou aqueles que se baseiam em supostas anomalias na ordem temporal da experiência consciente (Libet 1982, 1985). Argumentos dualistas de ambos os tipos são muito questionados por fisicalistas (P. S. Churchland 1981, Dennett e Kinsbourne 1992).

8.2. Teorias Fisicalistas

A maioria das outras teorias metafísicas da consciência é uma ou outra versão de algum tipo já conhecido de fisicalismo.

Teorias *eliminativistas* negam redutivamente a existência da consciência, ou ao menos a existência de alguns de seus tipos ou características comumente aceitas. (Ver o verbete sobre “o materialismo eliminativista”, por Ramsey 2011.) Os eliminativistas radicais rejeitam a própria noção de consciência como sendo confusa ou enganadora, e defendem que a distinção consciente/não-consciente não consegue cortar a realidade mental em suas articulações (Wilkes 1984, 1988). Eles entendem que a ideia de consciência é suficientemente equivocada para merecer ser eliminada e substituída por outros conceitos e distinções que melhor reflitam a verdadeira natureza da mente (P. S. Churchland 1983).

A maior parte dos eliminativistas é mais sofisticada em suas avaliações críticas. Ao invés de rejeitarem a noção completamente, questionam apenas algumas das características proeminentes que ela é comumente vista como envolvendo, tais como os qualia (Dennett 1990, Carruthers 2000), o eu consciente (Dennett 1992), ou o chamado “teatro cartesiano”, onde a sequência

temporal da experiência consciente é projetada internamente (Dennett e Kinsbourne 1992). Portanto, eliminativistas mais modestos, como Dennett, normalmente combinam suas negações mais qualificadas com uma teoria positiva daqueles aspectos da consciência que tomam como reais, tal como acontece no Modelo de Múltiplas Camadas [*Multiple Drafts Model*] (seção 9.3 a seguir).

A *teoria da identidade*, ao menos a teoria da identidade psicofísica tipo-tipo [*type-type*] estrita, oferece uma outra opção fortemente redutivista ao identificar propriedades, estados e processos mentais conscientes com físicos, normalmente de natureza neural ou neurofisiológica. Se ter uma experiência consciente qualitativa do vermelho fenomenal é *simplesmente* estar em um estado cerebral com as propriedades neurofisiológicas relevantes, então essas propriedades experienciais são reais, mas a realidade delas é uma realidade diretamente física.

A *teoria da identidade tipo-tipo* é chamada assim porque identifica tipos ou propriedades mentais e físicas, da mesma forma que se identifica a propriedade de ser água com a propriedade de ser composto de moléculas de H₂O. Após um breve período de popularidade nos primeiros dias do fisicalismo contemporâneo, nos anos 1950 e 60 (Place 1956, Smart 1959), ela tem sido bem menos amplamente adotada devido a problemas como a objeção da realização múltipla, de acordo com a qual propriedades mentais são mais abstratas e por isso capazes de serem realizadas por estruturas de fundo ou substratos químicos diferentes (Fodor 1974, Hellman e Thompson 1975). Se uma e a mesma propriedade consciente pode ser realizada por diferentes propriedades neurofisiológicas (ou mesmo não-neurofisiológicas) em diferentes organismos, então as duas propriedades não podem ser estritamente idênticas.

Entretanto, a teoria da identidade tipo-tipo gozou um renascimento recente, ainda que modesto, ao menos no que se refere aos qualia, ou propriedades conscientes qualitativas. Isso se deu, em parte, porque tratar a ligação psico-física como uma identidade foi visto por alguns como propiciando uma forma de dissolver o problema da lacuna explicativa (Hill e McLaughlin 1998, Papineau 1995, 2003). Eles argumentam que se a propriedade qualitativa consciente e a propriedade neural são idênticas, então não há necessidade de explicar como a última causa ou faz surgir a primeira. Uma não *causa* a outra: uma é a outra. Dessa forma, não há lacuna a ser preenchida, e nenhuma explicação adicional é necessária. Identidades não são o tipo de coisa que pode ser explicada, uma vez que nada é idêntico a nada além de si mesmo, e não faz

sentido perguntar porque algo é idêntico a si mesmo.

No entanto, outros contestam que o apelo à identidade tipo-tipo não elimina de forma tão óbvia a necessidade de uma explicação (Levine 2001). Mesmo quando duas descrições ou conceitos de fato se referem a uma e a mesma propriedade, pode-se ainda esperar, racionalmente, por alguma explicação dessa convergência, alguma explicação de como selecionaram uma e a mesma coisa, ainda que, inicialmente ou intuitivamente, não pareçam fazer isso. Em outros casos de identidades de propriedades empiricamente descobertas, tal como entre calor e energia cinética, há uma história a ser contada que explica a convergência correferencial, e parece justo esperar o mesmo no caso psicofísico. Portanto, apelar para identidades tipo-tipo talvez não seja suficiente, por si só, para dissolver o problema da lacuna explicativa.

A maior parte das teorias fiscalistas da consciência não são nem eliminativistas nem baseadas em identidades tipo-tipo estritas. Elas reconhecem a realidade da consciência, mas procuram situá-la no interior do mundo físico, por meio de alguma relação psicofísica mais fraca que a estrita identidade de propriedade.

Entre as variantes comuns, estão aquelas que tomam a realidade da consciência como *superveniente* sobre o físico, como *composta do* físico ou como sendo *realizada pelo* físico.

Teorias *funcionalistas*, em particular, apoiam-se largamente na noção de *realização* para explicar a relação entre a consciência e o físico. De acordo com o funcionalismo, um estado ou processo conta como sendo de um dado tipo mental ou consciente em virtude do papel funcional que executa no interior de um sistema apropriadamente organizado (Block 1980a). Um dado estado físico realiza o tipo mental consciente relevante ao executar o papel apropriado no interior de um sistema físico mais amplo que o contém. (Ver o verbete sobre o “funcionalismo”, por Levin 2010.) O funcionalista apela muitas vezes a analogias com outras relações entre níveis, como entre o biológico e o bioquímico, ou entre o químico e o atômico. Em cada um desses casos, propriedades e fatos de um certo nível são realizados por interações complexas entre itens de um nível subjacente.

Os críticos do funcionalismo negam frequentemente que a consciência possa ser adequadamente explicada em termos funcionais (Block 1980a, 1980b, Levine 1983, Chalmers 1996). De acordo com alguns críticos, a consciência pode ter características funcionais interessantes, mas sua natureza não é

essencialmente funcional. Essas afirmações são por vezes sustentadas por apelo à suposta possibilidade dos qualia ausentes ou invertidos, ou seja, a possibilidade de existirem seres funcionalmente equivalentes aos humanos normais, mas que possuem qualia invertidos ou não possuem nenhum qualia. O estatuto dessas possibilidades é controverso (Shoemaker 1981, Dennett 1990, Carruthers 2000), mas se elas são aceitas, parecem suscitar um problema para o funcionalista. (Ver o verbete “Qualia”, por Tye 2009.)

Aqueles que apoiam o fisicalismo ontológico na relação de realização frequentemente combinam isso com uma visão não-redutivista, no nível conceitual ou representacional, que ressalta a autonomia das ciências especiais e os modos distintos de descrição e de acesso cognitivo que elas ensejam.

O *fisicalismo não-redutivista* desse tipo nega que os recursos teóricos e conceituais apropriados e adequados para lidar com fatos do nível de substrato subjacente, ou nível da realização, tenham de ser igualmente adequados para lidar com aqueles do nível realizado (Putnam 1975, Boyd 1980). Como foi visto anteriormente, em resposta à pergunta sobre como a consciência pode existir, pode-se acreditar que todos os fatos econômicos são fisicamente realizados sem, com isso, pensar que os recursos das ciências físicas oferecem todas as ferramentas cognitivas e conceituais de que precisamos para fazer economia (Fodor 1974).

O fisicalismo não-redutivista tem sido criticado pela suposta incapacidade de “pagar sua dívida com o fisicalismo” em moeda redutivista. Seu erro é supostamente o de não dar uma explicação adequada de como propriedades conscientes são, ou poderiam ser, realizadas por estruturas e processos neurais, físicos ou funcionais subjacentes (Kim 1987, 1998). Ele foi, na verdade, acusado de incoerência, devido à sua tentativa de combinar a afirmação da realização física com a negação da capacidade de explicitar essa relação de uma forma estrita e inteligível *a priori* (Jackson 2004).

Entretanto, como observado antes na discussão sobre a questão de como a consciência existe, os fisicalistas não-redutivistas reagem concordando que uma abordagem da realização psicofísica é de fato necessária, mas adicionam que a abordagem relevante pode diferir muito de uma dedutibilidade *a priori* e, ainda assim, ser suficiente para satisfazer nossas legítimas demandas explicativas (McGinn 1991, Van Gulick 1985). A questão continua em discussão.

9. Teorias Específicas da Consciência

Apesar de existirem várias teorias metafísicas/ontológicas gerais da consciência, a lista de teorias específicas detalhadas acerca de sua natureza é ainda mais longa e mais diversificada. Nenhuma investigação breve poderia chegar perto de ser exaustiva, mas seis tipos principais de teorias podem ajudar a indicar o leque básico de opções: teorias de ordem superior [*higher-order theories*], teorias representacionais, teorias cognitivas, teorias neurais, teorias quânticas e teorias não-físicas.

9.1. Teorias de Ordem Superior

Teorias de ordem superior (OS) ²⁶ analisam a noção de estado mental consciente em termos de autoconsciência metamental reflexiva. A ideia central é que o que torna um estado mental M um estado mental consciente é o fato de estar acompanhado de um estado simultâneo e não-inferencial de ordem superior (isto é, metamental), cujo conteúdo é que se está no estado M no momento. Ter um desejo consciente por chocolate envolve ter dois estados mentais: é preciso ter tanto um desejo por chocolate, quanto um estado de ordem superior cujo conteúdo é que se está, no momento, tendo exatamente esse desejo. Estados mentais inconscientes são inconscientes precisamente por não termos os estados de ordem superior relevantes sobre eles. Seu caráter inconsciente consiste no fato de não estarmos reflexiva e diretamente cientes de estarmos em tais estados. (Ver o verbete sobre “teorias de ordem superior da consciência”, por Carruthers 2011.)

Teorias de ordem superior aparecem em duas variantes principais, que diferem quanto ao modo psicológico dos estados mentais constitutivos da consciência relevante. Teorias do pensamento de ordem superior (PenOS) tomam o estado de ordem superior requerido como sendo um metaestado assertórico similar ao pensamento (Rosenthal 1986, 1993). Teorias da percepção de ordem superior (PerOS) os veem mais como estados perceptivos, associados a uma espécie de sentido interno e a sistemas de monitoramento intramentais de algum tipo (Armstrong 1981, Lycan 1987, 1996).²⁷

²⁶ A sigla original é “HO”, de “*higher-order*”.

²⁷ As siglas originais são, respectivamente, “HOT” (*Higher-Order Thought*) e “HOP” (*Higher-Order Perception*).

Cada uma tem suas relativas vantagens e problemas. Proponentes da PenOS ressaltam que não temos nenhum órgão de sentido interno, e afirmam que não experienciamos nenhuma qualidade sensível a não ser aquelas que nos são apresentadas pela percepção direcionada ao exterior. Proponentes da PerOS, por outro lado, podem argumentar que sua abordagem explica algumas das condições adicionais exigidas por teorias OS como sendo consequências naturais da natureza perceptiva dos estados de ordem superior relevantes. Em particular, a exigência de que o metaestado constitutivo da consciência seja não-inferencial e simultâneo ao objeto mental de nível inferior pode ser explicada pelas condições correspondentes que normalmente se aplicam à percepção. Percebemos o que está acontecendo agora, e o fazemos de uma forma que não envolve nenhuma inferência, ao menos não uma inferência explícita no nível pessoal. Essas condições são igualmente necessárias para a PenOS, mas não são explicadas por essa teoria, o que aparentemente dá uma vantagem explicativa ao modelo PerOS (Lycan 2004, Van Gulick 2000), ainda que alguns teóricos da PerOS argumentem contrariamente (Carruthers 2000).

Sejam quais forem seus respectivos méritos, tanto teorias PerOS quanto PenOS enfrentam alguns desafios comuns, incluindo o que poderia ser chamado de *problema da generalidade*. Ter um pensamento ou uma percepção de um dado item *X* – seja uma pedra, uma caneta ou uma batata – não faz, em geral, de *X* um *X* consciente. Ver uma batata em cima do balcão, ou pensar nela, não faz dela uma batata consciente. Por que deveria, então, o fato de ter um pensamento ou uma percepção de um determinado desejo ou memória tornar esse desejo ou essa memória consciente? (Dretske 1995, Byrne 1997). Também não basta ressaltar que não utilizamos o termo “consciente” para pedras ou canetas que percebemos ou pensamos, mas apenas para estados mentais que percebemos ou a respeito dos quais pensamos (Lycan 1997, Rosenthal 1997). Isso pode bem ser verdade, mas o que falta é uma explicação de por que isso é adequado.

A abordagem de ordem superior é mais claramente relevante para as formas metamentais de consciência, mas alguns de seus defensores acreditam que ela explica também outros tipos de consciência, incluindo os tipos mais subjetivos de consciência como-é e qualitativa. Uma estratégia comum é analisar os qualia como características mentais que são capazes de ocorrer inconscientemente; por exemplo, eles podem ser explicados como propriedades de estados internos, cujas relações estruturadas de semelhança fazem surgir crenças sobre semelhanças objetivas no mundo (Shoemaker 1975, 1990).

Apesar dos qualia inconscientes poderem desempenhar o papel funcional, não precisa haver um como-é estar num estado que os possua (Nelkin 1989, Rosenthal 1991, 1997). De acordo com o teórico da OS, o caráter *como-é* [*what-it's-likeness*] surge apenas quando nos tornamos cientes de um estado de primeira ordem e de suas propriedades qualitativas através de um metaestado apropriado direcionado a ele.

Críticos da abordagem OS questionaram essa perspectiva, e alguns argumentaram que a noção de qualia inconsciente, na qual se baseia, é incoerente (Papineau 2002). Independentemente de quão satisfatória seja a abordagem dos qualia proposta pelas teorias OS, é importante ressaltar que a maior parte dos defensores da OS acreditam estar oferecendo uma teoria ampla da consciência, ou ao menos o cerne de uma teoria geral desse tipo, ao invés de simplesmente uma teoria restrita a alguns tipos metamentais especiais de consciência.

Outras variantes da teoria OS vão além das tradicionais versões PenOS e PerOS, incluindo aquelas que analisam a consciência em termos de pensamentos disposicionais de ordem superior, ao invés de pensamentos ocorrentes (Carruthers 2000). Outras apelam para um entendimento de ordem superior implícito, ao invés de explícito, e enfraquecem ou eliminam a pressuposição tradicional de que um metaestado precisa ser distinto e separado de seus objetos de ordem inferior (Gennaro 1995, Van Gulick 2000, 2004). Outras variantes da teoria OS continuam sendo oferecidas, e o debate entre defensores e críticos da abordagem básica continua vivo (Ver os artigos recentes em Gennaro 2004).

9.2. Teorias Representacionistas

Praticamente todas as teorias da consciência a veem como possuindo características representacionais, mas as assim chamadas teorias representacionistas se definem pela afirmação mais forte de que suas características representacionais exaurem suas características mentais (Harman 1990, Tye 1995, 2000). De acordo com o representacionista, estados mentais conscientes não possuem nenhuma propriedade mental além de suas propriedades representacionais. Portanto, dois estados conscientes ou experienciais que compartilham todas as suas propriedades representacionais não diferirão em nenhum aspecto mental.

A força exata dessa afirmação depende de como se interpreta a noção

de ser “representacionalmente o mesmo”, havendo várias opções plausíveis de critérios. Pode-se defini-la, grosso modo, em termos de condições de satisfação ou de verdade, mas, assim entendida, a tese representacionista parece claramente falsa. Existem várias maneiras em que estados podem compartilhar suas condições de satisfação ou verdade e, ainda assim, diferirem mentalmente, incluindo aqueles referentes aos modos de conceitualizar ou apresentar essas condições.

No extremo oposto, pode-se tomar dois estados como representacionalmente distintos se diferirem em qualquer característica que desempenhe um papel na operação ou função representacional deles. Numa interpretação liberal desse tipo, qualquer diferença nos portadores do conteúdo conta como diferença representacional, ainda que tenham o mesmo conteúdo intencional ou representacional, podem diferir apenas em seus *meios* ou *modos* de representação, e não em seus *conteúdos*.

Evidentemente, essa interpretação aumenta a plausibilidade da afirmação de que as propriedades representacionais de um estado consciente exaurem suas propriedades mentais, mas ao custo de enfraquecer significativamente, ou mesmo trivializar, a tese. Portanto, o representacionista parece precisar de uma interpretação da *igualdade representacional* que vá além das meras condições de satisfação, e que reflita todos os aspectos intencionais e relativos ao conteúdo da representação, sem ser sensível a meras diferenças em características subjacentes sem conteúdo dos processos no nível de realização. Portanto, a maior parte dos representacionistas estipulam condições para experiências conscientes que incluem, ao lado de uma condição relativa ao conteúdo, algum papel causal, ou exigências de configuração adicionais (Tye 1995, Dretske 1995, Carruthers 2000). Outros representacionistas aceitam a existência dos qualia, mas tratam-nos como propriedades objetivas que objetos externos são representados como possuindo. Ou seja, tratam-nos como *propriedades representadas*, ao invés de *propriedades de representações* ou de estados mentais (Dretske 1995, Lycan 1996).

O representacionismo pode ser visto como uma forma qualificada de eliminativismo, uma vez que nega a existência de propriedades de um tipo que estados mentais conscientes são usualmente pensados como possuindo – ou ao menos como parecendo possuir – a saber, aquelas que são mentais, mas não representacionais. Qualia, ao menos se entendidos como propriedades monádicas intrínsecas de estados conscientes acessíveis à introspecção, pareceriam ser os alvos mais óbvios dessa eliminação. De fato, parte da

motivação do representacionalismo é mostrar que podemos acomodar todos os fatos sobre a consciência, talvez dentro de um quadro fiscalista, sem precisarmos encontrar um lugar para os qualia ou para qualquer outra suposta propriedade mental não-representacional (Dennett 1990, Lycan 1996, Carruthers 2000).

O representacionalismo tem sido muito popular nos últimos anos e possui muitos defensores, mas continua altamente controverso e contraintuitivo quanto a casos e experimentos mentais cruciais (Block 1996). Em particular, a possibilidade de qualia invertidos fornece um caso crucial para teste. Para antirrepresentacionalistas, a mera possibilidade lógica de qualia invertidos mostra que estados conscientes podem diferir em um aspecto mental significante e, ao mesmo tempo, coincidir representacionalmente. Representacionalistas, em resposta, ou negam a possibilidade de tal inversão, ou sua alegada significação (Dretske 1995, Tye 2000).

Muitos outros argumentos foram levantados a favor ou contra o representacionalismo, tais como aqueles a respeito das percepções em diferentes modalidades sensoriais do mesmo estado de coisas – ver e tocar o mesmo cubo -, o que parece envolver diferenças mentais distintas de como os estados relevantes representam o mundo como sendo (Peacocke 1983, Tye 2003). Em cada caso, ambos os lados podem reunir fortes intuições e engenho argumentativo. Um vívido debate continua.

9.3. Teorias Cognitivas

Modelos que procuram explicar a consciência em termos de processos cognitivos têm sido oferecidos tanto por filósofos quanto por psicólogos.

O exemplo filosófico de maior destaque é o Modelo de Múltiplas Camadas (MMC) da consciência, proposto por Daniel Dennett (1991)²⁸. Ele combina elementos tanto do representacionalismo, quanto da teoria de ordem superior, mas o faz de tal maneira que o distancia, de forma interessante, das versões convencionais de ambos. O MMC inclui muitas características distintas, mas inter-relacionadas.

Seu nome reflete o fato de que, num momento qualquer, determinações de conteúdos de vários tipos ocorrem em todo o cérebro. O que faz alguns desses conteúdos serem conscientes não é o fato de ocorrerem num local

²⁸ A sigla original é “MDM”, de “*Multiple Drafts Model*”.

espacial ou funcional privilegiado – o chamado “Teatro Cartesiano” –, nem o fato de ocorrerem de um modo ou num formato especial: tudo isso é negado pelo MMC. Ao contrário, isso é uma questão do que Dennett chama de “celebridade cerebral” [*cerebral celebrity*], isto é, o grau no qual um dado conteúdo influencia o desenvolvimento futuro de outros conteúdos por todo o cérebro, especialmente no que concerne como esses efeitos se manifestam nos relatos e comportamentos da pessoa em resposta a várias indagações que podem indicar seu estado consciente.

Uma das afirmações principais do MMC é que indagações diferentes (por exemplo, receber diferentes questões ou estar em contextos diferentes que fazem demandas comportamentais diferentes) podem levar a respostas diferentes acerca do estado consciente da pessoa. Ademais, de acordo com o MMC, pode ser que não haja uma questão de fato, independente da investigação, sobre qual era, efetivamente, o estado consciente da pessoa. Daí o “múltiplo” de Modelos de Múltiplas Camadas.

O MMC é representacionalista, pois analisa a consciência em termos de relações de conteúdo. Ele também nega a existência dos qualia e, portanto, rejeita qualquer tentativa de distinguir estados conscientes de estados não-conscientes por meio de sua presença. Rejeita, também, a noção do eu como um observador interno, esteja ele localizado no Teatro Cartesiano ou em qualquer outro lugar. O MMC trata o eu como um aspecto emergente ou virtual da narrativa aproximadamente serial e coerente que é construída pelo jogo interativo dos conteúdos no sistema. Muitos desses conteúdos estão interligados no nível intencional, como percepções ou fixações de um ponto de vista relativamente unificado e estendido temporalmente, ou seja, são coerentes em seus conteúdos como se fossem as experiências de um eu em movimento. Mas é a ordem de dependência que é crucial para a abordagem de MMC. Os conteúdos relevantes não são unificados porque são todos observados por um único eu, mas exatamente o contrário. É porque são unificados e coerentes no nível do conteúdo que eles contam como experiências de um eu único, ao menos de um eu virtual único.

É nesse aspecto que o MMC compartilha alguns elementos com as teorias de ordem superior. Os conteúdos que compõem a narrativa em série são, ao menos implicitamente, aqueles de um eu em fluxo, mesmo que virtual, e são eles que são mais provavelmente expressos nos relatos que a pessoa faz de seu estado consciente, em resposta às várias indagações. Eles envolvem, então, um certo grau de reflexividade ou autoconsciência de um tipo que é central para as

teorias de ordem superior, mas o aspecto de ordem superior é mais uma característica implícita da corrente de conteúdos do que presente em estados explícitos e distintos de ordem superior, do tipo encontrado em teorias OS convencionais.

O MMC de Dennett é altamente influente, mas suscitou também críticas, principalmente por parte daqueles que o acharam insuficientemente realista na sua abordagem da consciência e, na melhor das hipóteses, incompleto quanto à execução de seu proclamado objetivo de explicá-la completamente (Block 1994, Dretske 1994, Levine 1994). Muitos de seus críticos reconhecem o valor e *insight* do MMC, mas negam que não haja outros fatos reais sobre a consciência além daqueles capturados pela teoria (Rosenthal 1994, Van Gulick 1994, Akins 1996).

Um grande exemplo psicológico da abordagem cognitiva é a teoria do ambiente de trabalho global, de Bernard Baars (1988). Ela oferece um modelo psicológico e, em grande medida, funcional da consciência, que lida mais diretamente com a noção de consciência de acesso e tem muito em comum com a ideia da consciência como uma forma de celebridade cerebral, do modelo de múltiplas camadas. A principal ideia das teorias de ambiente de trabalho global é que a consciência é uma capacidade ou módulo de recurso limitado que disponibiliza a informação a ser “transmitida” amplamente por todo o sistema, e que permite um processamento sofisticado mais flexível. Ela é, assim, bastante similar aos vários modelos em psicologia cognitiva sobre a atenção e a memória de trabalho. Muitos outros modelos cognitivos da consciência baseados psicologicamente foram propostos, incluindo aqueles de George Mandler (1975), Timothy Shallice (1988), e Daniel Schacter (1989).

9.4. Teorias Neurais

Teorias neurais da consciência assumem várias formas, ainda que de alguma maneira a maioria se refira aos chamados “correlatos neurais da consciência”, ou CNC²⁹. A menos que se seja um dualista ou um outro tipo de não-fisicalista, algo mais do que mera correlação é necessário; ao menos alguns CNC precisam ser os substratos essenciais da consciência. Uma teoria neural informativa precisa explicar por que ou como as correlações relevantes existem, e se a teoria está comprometida com o fisicalismo, precisará mostrar como os substratos neurais subjacentes podem ser idênticos a estados conscientes, ou

²⁹ A sigla original é “NCCs”, de “*Neural Correlates of Consciousness*”.

ao menos realizá-los ao executar suas funções ou condições exigidas (Metzinger 2000).

Essas teorias são diversificadas, não apenas quanto aos processos ou propriedades neurais a que apelam, mas também quanto aos aspectos da consciência que tomam como seus respectivos *explananda*. Algumas são baseadas em características sistêmicas de alto nível do cérebro, mas outras se concentram em propriedades fisiológicas ou estruturais mais específicas, com as correspondentes diferenças em suas pretendidas metas explicativas.

Uma amostra de teorias neurais recentes deveria incluir modelos que apelam para campos integrados globais (Kinsbourne), ligações através de oscilações sincrônicas (Singer 1999, Crick e Koch 1990), conjunções neurais transitórias mediadas por NMDA (Flohr 1995), padrões talâmicos modulados de ativação cortical (Llinas 2001), circuitos corticais de reentrada (Edelman 1989), mecanismos comparativos que executam círculos contínuos de ação-predição-avaliação entre as áreas frontais e o mesencéfalo (Gray 1995), hemisfério esquerdo baseado em processos interpretativos (Gazzaniga 1988), e processos homeostáticos somatossensórios emotivos baseados nonexo frontal-límbico (Damásio 1999), ou na substância cinzenta periaquedutal (Panksepp 1998).

Em cada caso, o objetivo é explicar como a organização e a atividade no nível neural relevante poderiam estar por trás de uma ou outra característica ou tipo importante de consciência. Campos globais ou conjunções sincrônicas transitórias poderiam estar por trás da unidade intencional da consciência fenomenal. Plasticidade baseada em NMDA, projeções talâmicas específicas no córtex ou ondas oscilatórias regulares poderiam todas contribuir para a formação de padrões ou regularidades neurais de curto prazo, mas bem difundidas, necessárias para unir experiências conscientes integradas fora da atividade local, em diversos módulos cerebrais especializados. Processos interpretativos do hemisfério esquerdo poderiam oferecer uma base para formas narrativas de autoconsciência. Assim, é possível que múltiplas teorias neurais distintas sejam todas verdadeiras, cada uma contribuindo com algum entendimento parcial das ligações entre a mentalidade consciente, em suas diversas formas, e o cérebro ativo, em seus muitos níveis de organização e estrutura complexas.

9.5. Teorias Quânticas

Outras teorias físicas foram além do neural e situaram o lugar natural da consciência num nível bem mais fundamental: em particular, no nível micro-físico

dos fenômenos quânticos. De acordo com essas teorias, a natureza e a base da consciência não podem ser adequadamente compreendidas dentro do quadro da física clássica, mas devem ser buscadas dentro do quadro alternativo da realidade física fornecido pela mecânica quântica. Os proponentes da abordagem da consciência quântica veem a natureza radicalmente alternativa e muitas vezes contra-intuitiva da física quântica como exatamente o que é preciso para superar os supostos obstáculos explicativos que confrontam tentativas mais convencionais de eliminar a lacuna psico-física.

Mais uma vez, há um amplo leque de teorias e modelos específicos que foram propostos apelando para vários fenômenos quânticos, de forma a explicar várias características da consciência. Seria impossível catalogá-los aqui e mesmo explicar, de uma maneira substancial, as principais características da mecânica quântica às quais apelam. Entretanto, um breve exame seletivo pode prover um sentido, ainda que parcial e obscuro, das opções que foram propostas.

O físico Roger Penrose (1989, 1994) e o anestesiólogo Stuart Hameroff (1998) encabeçaram um modelo segundo o qual a consciência surge através de efeitos quânticos que ocorrem dentro de estruturas sub-celulares internas aos neurônios, chamadas de *microtúbulos*. O modelo postula os chamados “*colapsos objetivos*”, que se referem ao sistema quântico mudando de uma superposição de múltiplos estados possíveis para um único estado definido, mas sem a intervenção de um observador ou de uma medida, como na maior parte dos modelos mecânico-quânticos. De acordo com Penrose e Hameroff, o ambiente interno aos microtúbulos é especialmente apropriado para tais colapsos objetivos, e os auto-colapsos resultantes produzem um fluxo coerente que regula a atividade neuronal e torna possível processos mentais não-algorítmicos.

O psiquiatra Ian Marshall propôs um modelo que pretende explicar a unidade coerente da consciência apelando para a produção, no cérebro, de um estado físico parecido com um *condensado de Bose-Einstein*. Este é um fenômeno quântico no qual uma coleção de átomos age como uma única entidade coerente, e a distinção entre átomos discretos se perde. Apesar de estados cerebrais não serem, literalmente, exemplos de condensados de Bose-Einstein, razões foram dadas para mostrar por que cérebros são mais suscetíveis a gerar estados capazes de exibir uma coerência similar (Marshall e Zohar 1990).

Uma base para a consciência também foi buscada na natureza *holista*

da mecânica quântica e no fenômeno do *entrelaçamento*, de acordo com o qual partículas que interagiram continuam tendo suas naturezas dependentes uma da outra, mesmo depois da sua separação. Não surpreendentemente, esses modelos pretendem explicar principalmente a coerência da consciência, mas também foram invocados como um desafio mais geral à concepção atomista da física tradicional, segundo a qual as propriedades do todo são explicáveis por apelo às propriedades das partes mais seu modo de combinação, um método de explicação que hoje poderia ser visto como fracassado para explicar a consciência (Silberstein 1998, 2001).

Outros viram a mecânica quântica como indicando que a consciência é uma propriedade absolutamente fundamental da realidade física, uma propriedade que precisa ser introduzida no nível mais básico de todos (Stapp 1993). Apelaram principalmente para o papel do observador no colapso da função de onda, isto é, o colapso da realidade quântica de uma superposição de estados possíveis em um único estado definido, quando a medição é feita. Esses modelos podem ou não adotar uma forma de quase-idealismo, na qual a própria existência da realidade física depende dela ser conscientemente observada.

Existem muitos outros modelos quânticos da consciência que podem ser encontrados na literatura – alguns defendendo uma metafísica radicalmente revisionista, outros não –, mas esses quatro fornecem uma amostra razoável, ainda que parcial, das alternativas.

9.6. Teorias Não-Físicas

A maior parte das teorias específicas da consciência – sejam cognitivas, neurais ou da mecânica quântica – pretendem explicar ou propor modelos da consciência vista como uma característica natural do mundo físico. No entanto, aqueles que rejeitam uma ontologia fiscalista da consciência precisam encontrar formas alternativas de tratá-la como um aspecto não-físico da realidade. Portanto, aqueles que adotam uma visão metafísica dualista ou antifiscalista precisam, por fim, fornecer modelos específicos da consciência, diferentes dos cinco tipos mencionados. Tanto os dualistas de substância quanto os dualistas de propriedades precisam desenvolver os detalhes de suas teorias de forma a articular as naturezas específicas das características não-físicas relevantes da natureza, com as quais igualam a consciência ou às quais apelam para explicá-la.

Uma grande variedade desses modelos foi proposta, incluindo os

seguintes. David Chalmers (1996) ofereceu uma versão admitidamente especulativa de pampsiquismo, que apela para a noção de informação, não apenas para explicar invariâncias psicofísicas entre o fenomenal e espaços informacionais fisicamente realizados, mas possivelmente também para explicar a ontologia do físico como em si mesma derivada do informacional (uma versão da teoria do “*it from bit*”). De forma um tanto similar, Gregg Rosenberg propôs recentemente (2004) uma abordagem da consciência que, simultaneamente, trata a base categorial última das relações causais. Tanto no caso causal quanto no caso consciente, Rosenberg argumenta que fatos funcionais e relacionais precisam, no fim das contas, depender de uma base categorial não-relacional, e oferece um modelo segundo o qual relações causais e fatos fenomenais qualitativos ambos dependem da mesma base. Ademais, como mencionado anteriormente (seção 9.5), algumas teorias quânticas tratam a consciência como uma característica fundamental da realidade (Stapp 1993) e, na medida em que assim o fazem, também poderiam ser plausivelmente classificadas como teorias não-físicas.

10. Conclusão

Uma compreensão completa da consciência provavelmente exigirá teorias de vários tipos. Pode-se, com utilidade e sem contradição, aceitar uma diversidade de modelos que, cada um à sua maneira, buscam explicar, respectivamente, os aspectos físicos, neurais, cognitivos, funcionais, representacionais e de ordem superior da consciência. É improvável que haja uma perspectiva teórica única que seja suficiente para explicar todas as características da consciência que pretendemos compreender. Portanto, uma abordagem sintética e pluralista deve oferecer a melhor via para futuros progressos.

Referências

- Akins, K. 1993. “A bat without qualities?” In M. Davies and G. Humphreys, eds. *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Akins, K. 1996. “Lost the plot? Reconstructing Dennett’s multiple drafts theory of consciousness.” *Mind and Language*, 11: 1-43.
- Anderson, J. 1983. *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Armstrong, D. 1968. *A Materialist Theory of Mind*, London: Routledge and Kegan

Paul.

Armstrong, D. 1981. "What is consciousness?" In *The Nature of Mind*. Ithaca, NY: Cornell University Press.

Baars, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Balog, K. 1999. "Conceivability, possibility, and the mind-body problem." *Philosophical Review*, 108: 497-528.

Block, N. 1980a. "Troubles with Functionalism," in *Readings in the Philosophy of Psychology*, Volume 1, Ned Block, ed., Cambridge, MA : Harvard University Press, 268-305.

Block, N. 1980b. "Are absent qualia impossible?" *Philosophical Review*, 89/2: 257-74.

Block, N. 1990. "Inverted Earth," *Philosophical Perspectives*, 4, J. Tomberlin, ed., Atascadero, CA: Ridgeview Publishing Company.

Block, N. 1995. "On a confusion about the function of consciousness." *Behavioral and Brain Sciences*, 18: 227-47.

Block, N. 1994. "What is Dennett's theory a theory of?" *Philosophical Topics*, 22/1-2: 23-40.

Block, N. 1996. "Mental paint and mental latex." In E. Villanueva, ed. *Perception*. Atascadero, CA: Ridgeview.

Block, N. and Stalnaker, R. 1999. "Conceptual analysis, dualism, and the explanatory gap." *Philosophical Review*, 108/1: 1-46.

Boyd, R. 1980. "Materialism without reductionism: What physicalism does not entail." In N.

Block, ed. *Readings in the Philosophy of Psychology*, Vol. 1. Cambridge, MA: Harvard University Press.

Brook, A. and Raymont, P. 2010. "The Unity of Consciousness", *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Edward N. Zalta (ed.), URL =<<http://plato.stanford.edu/archives/fall2010/entries/consciousness-unity/>>.

Byrne, A. 1997. "Some like it HOT: consciousness and higher-order thoughts." *Philosophical Studies*, 2: 103-29.

Byrne, A. 2001. "Intentionalism defended". *Philosophical Review*, 110: 199-240.

Campbell, K. 1970. *Body and Mind*. New York: Doubleday.

Campbell, J. 1994. *Past, Space, and Self*. Cambridge, MA: MIT Press.

- Carruthers, P. 2000. *Phenomenal Consciousness*. Cambridge: Cambridge University Press.
- Carruthers, P. 2011. "Higher-Order Theories of Consciousness", *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2011/entries/consciousness-higher/>>.
- Chalmers, D. 1995. "Facing up to the problem of consciousness". *Journal of Consciousness Studies*, 2: 200-19.
- Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Chalmers, D. 2002. "Does conceivability entail possibility?" In T. Gendler and J. Hawthorne eds. *Conceivability and Possibility*. Oxford: Oxford University Press.
- Chalmers, D. 2003. "The content and epistemology of phenomenal belief." In A. Jokic and Q. Smith eds. *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.
- Chalmers, D. and Jackson, F. 2001. "Conceptual analysis and reductive explanation". *Philosophical Review*, 110/3: 315-60.
- Churchland, P. M. 1985. "Reduction, qualia, and direct introspection of brain states". *Journal of Philosophy*, 82: 8-28.
- Churchland, P. M. 1995. *The Engine of Reason and Seat of the Soul*. Cambridge, MA: MIT Press.
- Churchland, P. S. 1981. "On the alleged backwards referral of experiences and its relevance to the mind body problem". *Philosophy of Science*, 48: 165-81.
- Churchland, P. S. 1983. "Consciousness: the transmutation of a concept". *Pacific Philosophical Quarterly*, 64: 80-95.
- Churchland, P. S. 1996. "The hornswoggle problem". *Journal of Consciousness Studies*, 3: 402-8.
- Clark, A. 1993. *Sensory Qualities*. Oxford: Oxford University Press.
- Clark, G. and Riel-Salvatore, J. 2001. "Grave markers, middle and early upper paleolithic burials". *Current Anthropology*, 42/4: 481-90.
- Cleeremans, A., ed. 2003. *The Unity of Consciousness: Binding, Integration and Dissociation*. Oxford: Oxford University Press.
- Crick, F. and Koch, C. 1990. "Toward a neurobiological theory of consciousness". *Seminars in Neuroscience*, 2: 263-75.
- Crick, F. H. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Scribners.

- Davies, M. and Humphreys, G. 1993. *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Damasio, A. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt.
- Dennett, D. C. 1978. *Brainstorms*. Cambridge: MIT Press.
- Dennett, D. C. 1984. *Elbow Room: The Varieties of Free Will Worth Having*. Cambridge: MIT Press.
- Dennett, D. C. 1990. "Quining qualia". In *Mind and Cognition*, W. Lycan, ed., Oxford : Blackwell, 519-548.
- Dennett, D. C. 1991. *Consciousness Explained*. Boston: Little, Brown and Company.
- Dennett, D. C. 1992. "The self as the center of narrative gravity". In F. Kessel, P. Cole, and D. L. Johnson, eds. *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Lawrence Erlbaum.
- Dennett, D. C. 2003. *Freedom Evolves*. New York: Viking.
- Dennett, D. C. and Kinsbourne, M. 1992. "Time and the observer: the where and when of consciousness in the brain". *Behavioral and Brain Sciences*, 15: 187-247.
- Descartes, R. 1644/1911. *The Principles of Philosophy*. Translated by E. Haldane and G. Ross. Cambridge: Cambridge University Press.
- Dretske, F. 1993. "Conscious experience." *Mind*, 102: 263-283.
- Dretske, F. 1994. "Differences that make no difference". *Philosophical Topics*, 22/1-2: 41-58.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, Mass: The MIT Press, Bradford Books.
- Eccles, J. and Popper, K. 1977. *The Self and Its Brain: An Argument for Interactionism*. Berlin: Springer
- Edelman, G. 1989. *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.
- Farah, M. 1990. *Visual Agnosia*. Cambridge: MIT Press.
- Flanagan, O. 1992. *Consciousness Reconsidered*. Cambridge, MA: MIT Press.
- Flohr, H. 1995. "An information processing theory of anesthesia". *Neuropsychologia*, 33/9: 1169-80.

- Flohr, H., Glade, U. and Motzko, D. 1998. "The role of the NMDA synapse in general anesthesia". *Toxicology Letters*, 100-101: 23-29.
- Fodor, J. 1974. "Special sciences". *Synthese*, 28: 77-115.
- Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Foster, J. 1989. "A defense of dualism". In J. Smythies and J. Beloff, eds. *The Case for Dualism*. Charlottesville, VA: University of Virginia Press.
- Foster J. 1996. *The Immaterial Self: A Defence of the Cartesian Dualist Conception of Mind*. London: Routledge.
- Gallistel, C. 1990. *The Organization of Learning*. Cambridge, MA: MIT Press.
- Gardiner, H. 1985. *The Mind's New Science*. New York: Basic Books.
- Gazzaniga, M. 1988. *Mind Matters: How Mind and Brain Interact to Create our Conscious Lives*. Boston: Houghton Mifflin.
- Gennaro, R. 1995. *Consciousness and Self-consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*. Amsterdam and Philadelphia: John Benjamins.
- Gennaro, R., ed. 2004. *Higher-Order Theories of Consciousness*. Amsterdam and Philadelphia: John Benjamins.
- Gertler, B. 2011. "Self-Knowledge", *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N.Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2011/entries/self-knowledge/>>.
- Gray, J. 1995. "The contents of consciousness: a neuropsychological conjecture". *Behavior and Brain Sciences*, 18/4: 659-722.
- Hameroff, S. 1998. "Quantum computation in brain microtubules? The Penrose-Hameroff "Orch OR" model of consciousness". *Philosophical Transactions Royal Society London*, A 356: 1869-96.
- Hardin, C. 1986. *Color for Philosophers*. Indianapolis: Hackett.
- Hardin, C. 1992. "Physiology, phenomenology, and Spinoza's true colors". In A. Beckermann,
- H. Flohr, and J. Kim, eds. *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. Berlin and New York: De Gruyter.
- Harman, G. 1990. "The intrinsic quality of experience". In J. Tomberlin, ed. *Philosophical Perspectives*, 4. Atascadero, CA: Ridgeview Publishing .
- Hartshorne, C. 1978. "Panpsychism: mind as sole reality". *Ultimate Reality and Meaning*, 1: 115-29.

- Hasker, W. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Heidegger, M. 1927/1962. *Being and Time (Sein und Zeit)*. Translated by J. Macquarrie and E. Robinson. New York: Harper and Row.
- Hellman, G. and Thompson, F. 1975. "Physicalism: ontology, determination and reduction". *Journal of Philosophy*, 72: 551-64.
- Hill, C. 1991. *Sensations: A Defense of Type Materialism*. Cambridge: Cambridge University Press.
- Hill, C. 1997. "Imaginability, conceivability, possibility, and the mind-body problem". *Philosophical Studies*, 87: 61-85.
- Hill, C. and McLaughlin, B. 1998. "There are fewer things in reality than are dreamt of in Chalmers' philosophy". *Philosophy and Phenomenological Research*, 59/2: 445-54.
- Horgan, T. 1984. "Jackson on Physical Information and Qualia." *Philosophical Quarterly*, 34: 147-83.
- Horgan, T. and Tienson, J. 2002. "The intentionality of phenomenology and the phenomenology of intentionality". In D. J. Chalmers, ed., *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press.
- Hume, D. 1739/1888. *A Treatise of Human Nature*. ed. L Selby-Bigge. Oxford: Oxford University Press.
- Humphreys, N. 1982. *Consciousness Regained*. Oxford: Oxford University Press.
- Humphreys, N. 1992. *A History of the Mind*. London: Chatto and Windus.
- Husserl, E. 1913/1931. *Ideas: General Introduction to Pure Phenomenology (Ideen au einer reinen Phänomenologie und phänomenologischen Philosophie)*. Translated by W. Boyce Gibson. New York: MacMillan.
- Husserl, E. 1929/1960. *Cartesian Meditations: an Introduction to Phenomenology*. Translated by Dorian Cairns. The Hague: M. Nijhoff.
- Huxley, T. 1866. *Lessons on Elementary Physiology* 8. London
- Huxley, T. 1874. "On the hypothesis that animals are automata". *Fortnightly Review*, 95: 555-80. Reprinted in *Collected Essays*. London, 1893.
- Hurley, S. 1998. *Consciousness in Action*. Cambridge, MA: Harvard University Press.
- Jackson, F. 1982. "Epiphenomenal qualia". *Philosophical Quarterly*, 32: 127-136.
- Jackson, F. 1986. "What Mary didn't know". *Journal of Philosophy*, 83: 291-5.
- Jackson, F. 1993. "Armchair metaphysics". In J. O'Leary-Hawthorne and M.

- Michael, eds. *Philosophy of Mind*. Dordrecht: Kluwer Books.
- Jackson, F. 1998. "Postscript on qualia". In F. Jackson *Mind, Method and Conditionals*. London: Routledge.
- Jackson, F. 2004. "Mind and illusion." In P. Ludlow, Y. Nagasawa and D. Stoljar eds. *There's Something about Mary: Essays on the Knowledge Argument*. Cambridge, MA: MIT Press.
- James, W. 1890. *The Principles of Psychology*. New York: Henry Holt and Company.
- Jaynes, J. 1974. *The Origins of Consciousness in the Breakdown of the Bicameral Mind*. Boston: Houghton Mifflin.
- Kim, J. 1987. "The myth of non-reductive physicalism". *Proceedings and Addresses of the American Philosophical Association*.
- Kim, J. 1998. *Mind in Physical World*. Cambridge: MIT Press.
- Kirk, R. 2012. "Zombies", *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition), Edward N. Zalta (ed.), forthcoming URL=<http://plato.stanford.edu/archives/spr2012/entries/zombies/>.
- Kant, I. 1787/1929. *Critique of Pure Reason*. Translated by N. Kemp Smith. New York: MacMillan.
- Kinsbourne, M. 1988. Integrated field theory of consciousness". In A. Marcel and E. Bisiach, eds. *Consciousness in Contemporary Science*. Oxford: Oxford University Press.
- Kirk, R. 1974. "Zombies vs materialists". *Proceedings of the Aristotelian Society*, Supplementary Volume, 48: 135-52.
- Kirk, R. 1991. "Why shouldn't we be able to solve the mind-body problem?" *Analysis*, 51: 1723.
- Klima, G. 2008. "The Medieval Problem of Universals", *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2008/entries/universals-medieval/>.
- Köhler, W. 1929. *Gestalt Psychology*. New York: Liveright.
- Köffka, K. 1935. *Principles of Gestalt Psychology*. New York: Harcourt Brace.
- Leibniz, G. W. 1686 /1991. *Discourse on Metaphysics*. Translated by D. Garter and R. Aries. Indianapolis: Hackett.
- Leibniz, G. W. 1720/1925. *The Monadology*. Translated by R. Lotte. London: Oxford University Press.

- Levin, J. 2010. "Functionalism", *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.), URL=<<http://plato.stanford.edu/archives/sum2010/entries/functionalism/>>.
- Levine, J. 1983. "Materialism and qualia: the explanatory gap". *Pacific Philosophical Quarterly*, 64: 354-361.
- Levine, J. 1993. "On leaving out what it's like". In M. Davies and G. Humphreys, eds. *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Levine, J. 1994. "Out of the closet: a qualophile confronts qualophobia". *Philosophical Topics*, 22/1-2: 107-26.
- Levine, J. 2001. *Purple Haze: The Puzzle of Conscious Experience*. Cambridge, Mass: The MIT Press.
- Lewis, D. 1972. "Psychophysical and theoretical identifications". *Australasian Journal of Philosophy*, 50: 249-58.
- Lewis, D. 1990. "What experience teaches." In W. Lycan, ed. *Mind and Cognition: A Reader*. Oxford: Blackwell.
- Libet, B. 1982. "Subjective antedating of a sensory experience and mind-brain theories". *Journal of Theoretical Biology*, 114: 563-70.
- Libet, B. 1985. "Unconscious cerebral initiative and the role of conscious will in voluntary action". *Behavioral and Brain Sciences*, 8: 529-66.
- Llinas, R. 2001. *I of the vortex: from neurons to self*. Cambridge, MA: MIT Press
- Loar, B. 1990. "Phenomenal states," in *Philosophical Perspectives*, 4: 81-108.
- Loar, B. 1997. "Phenomenal states". In N. Block, O. Flanagan, and G. Guzeldere eds. *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Locke, J. 1688/1959. *An Essay on Human Understanding*. New York: Dover.
- Lockwood, M. 1989. *Mind, Brain, and the Quantum*. Oxford: Oxford University Press.
- Lorenz, K. 1977. *Behind the Mirror (Rückseite dyes Speigels)*. Translated by R. Taylor. New York: Harcourt Brace Jovanovich.
- Lycan, W. 1987. *Consciousness*. Cambridge, MA: MIT Press.
- Lycan, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Lycan, W. 2004. "The superiority of HOP to HOT". In R. Gennaro ed. *Higher-Order Theories of Consciousness*. Amsterdam and Philadelphia: John Benjamins.
- Lycan, W. 2008. "Representational Theories of Consciousness", *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed.), URL

= <<http://plato.stanford.edu/archives/fall2008/entries/consciousness/representational/>>.

Mandler, G. 1975. *Mind and Emotion*. New York: Wiley.

Marshall, I. and Zohar, D. 1990. *The Quantum Self: Human Nature and Consciousness Defined by the New Physics*. New York: Morrow.

McGinn, C. 1989. "Can we solve the mind-body problem?" *Mind*, 98: 349-66

McGinn, C. 1991. *The Problem of Consciousness*. Oxford: Blackwell.

McGinn, C. 1995. "Consciousness and space." In T. Metzinger, ed. *Conscious Experience*. Paderborn: Ferdinand Schöningh.

Merleau-Ponty, M. 1945/1962. *Phenomenology of Perception (Phénoménologie de l'ye Perception)*. Translated by Colin Smith. London: Routledge and Kegan Paul.

Metzinger, T., ed. 1995. *Conscious Experience*. Paderborn: Ferdinand Schöningh.

Metzinger, T. ed. 2000. *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.

Mill, J. 1829. *Analysis of the Phenomena of the Human Mind*. London.

Mill, J.S. 1865. *An Analysis of Sir William Hamilton's Philosophy*. London.

Moore, G. E. 1922. "The refutation of idealism." In G. E. Moore *Philosophical Studies*. London : Routledge and Kegan Paul.

Nagel, T. 1974. "What is it like to be a bat?" *Philosophical Review*, 83: 435-456.

Nagel, T. 1979. "Panpsychism." In T. Nagel *Mortal Questions*. Cambridge: Cambridge University Press.

Natsoulas, T. 1983. "Concepts of consciousness." *Journal of Mind and Behavior*, 4: 195-232.

Nelkin, N. 1989. "Unconscious sensations." *Philosophical Psychology*, 2: 129-41.

Nemirow, L. 1990. "Physicalism and the cognitive role of acquaintance." In W. Lycan, ed. *Mind and Cognition: A Reader*. Oxford: Blackwell.

Neisser, U. 1965. *Cognitive Psychology*. Englewood Cliffs: Prentice Hall.

Nida-Rümelin, M. 1995. "What Mary couldn't know: belief about phenomenal states." In T. Metzinger, ed. *Conscious Experience*. Paderborn: Ferdinand Schöningh.

Nida-Rümelin, M. 2010. "Qualia: The Knowledge Argument", *The Stanford Encyclopedia of Philosophy (Summer2010 Edition)*, Edward N. Zalta (ed.), URL =

- <<http://plato.stanford.edu/archives/sum2010/entries/qualia-knowledge/>>.
- O'Connor, T. 2011. "Free Will", *The Stanford Encyclopedia of Philosophy (Summer 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2011/entries/freewill/>>.
- Panksepp, J. 1998. *Affective Neuroscience*. Oxford: Oxford University Press.
- Papineau, D. 1994. *Philosophical Naturalism*. Oxford: Blackwell.
- Papineau, D. 1995. "The antipathetic fallacy and the boundaries of consciousness." In T. Metzinger, ed. *Conscious Experience*. Paderborn: Ferdinand Schöningh.
- Papineau, D. 2002. *Thinking about Consciousness*. Oxford: Oxford University Press.
- Peacocke, C. 1983. *Sense and Content*, Oxford: Oxford University Press.
- Pearson, M.P. 1999. *The Archeology of Death and Burial*. College Station, Texas: Texas A&M Press.
- Penfield, W. 1975. *The Mystery of the Mind: a Critical Study of Consciousness and the Human Brain*. Princeton, NJ: Princeton University Press.
- Perry, J. 2001. *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.
- Penrose, R. 1989. *The Emperor's New Mind: Computers, Minds and the Laws of Physics*. Oxford: Oxford University Press.
- Penrose, R. 1994. *Shadows of the Mind*. Oxford: Oxford University Press.
- Place, U. T. 1956. "Is consciousness a brain process?" *British Journal of Psychology*, 44-50.
- Putnam, H. 1975. "Philosophy and our mental life." In H. Putnam *Mind Language and Reality: Philosophical Papers Vol. 2*. Cambridge: Cambridge University Press.
- Putnam, H. and Oppenheim, P. 1958. "Unity of science as a working hypothesis." In H. Fiegl, G. Maxwell, and M. Scriven eds. *Minnesota Studies in the Philosophy of Science II*. Minneapolis: University of Minnesota Press.
- Ramsey, W. 2011. "Eliminative Materialism", *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2011/entries/materialism-eliminative/>>.
- Rey, G. 1986. "A question about consciousness." In H. Otto and J. Tuedio, eds. *Perspectives on Mind*. Dordrecht: Kluwer.

- Robinson, H. 1982. *Matter and Sense: A Critique of Contemporary Materialism*. Cambridge: Cambridge University Press.
- Robinson, H. 2011. "Dualism", *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*, Edward N. Zalta (ed.) UR = <<http://plato.stanford.edu/archives/win2011/entries/dualism/>>.
- Robinson, D. 1993. "Epiphenomenalism, laws, and properties." *Philosophical Studies*, 69: 1-34.
- Robinson, W. 2011. "Epiphenomenalism", *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2011/entries/epiphenomenalism/>>.
- Rosenberg, G. 2004. *A Place for Consciousness: Probing the Deep Structure of the Natural World*. New York: Oxford University Press.
- Rosenthal, D. 1986. "Two concepts of consciousness." *Philosophical Studies*, 49: 329-59.
- Rosenthal, D. 1991. "The independence of consciousness and sensory quality." In E. Villanueva, ed. *Consciousness*. Atascadero, CA: Ridgeview Publishing.
- Rosenthal, D. M. 1993. "Thinking that one thinks." In M. Davies and G. Humphreys, eds. *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Rosenthal, D. 1994. "First person operationalism and mental taxonomy." *Philosophical Topics*, 22/1-2: 319-50.
- Rosenthal, D. M. 1997. "A theory of consciousness." In N. Block, O. Flanagan, and G. Guzeldere, eds. *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Russell, B. 1927. *The Analysis of Matter*. London: Kegan Paul.
- Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson and Company.
- Sacks, O. 1985. *The Man who Mistook his Wife for a Hat*. New York: Summit.
- Seager, W. 1995. "Consciousness, information, and panpsychism." *Journal of Consciousness Studies*, 2: 272-88.
- Seager, W. e Allen-Hermanson, S. 2012. "Panpsychism", *The Stanford Encyclopedia of Philosophy (Spring 2012 Edition)*, Edward N. Zalta (ed.), forthcoming URL = <<http://plato.stanford.edu/archives/spr2012/entries/panpsychism/>>.
- Schacter, D. 1989. "On the relation between memory and consciousness:

- dissociable interactions and consciousness." In H. Roediger and F. Craik eds. *Varieties of Memory and Consciousness*. Hillsdale, NJ: Erlbaum.
- Schneider W. and Shiffrin, R. 1977. "Controlled and automatic processing: detection, search and attention." *Psychological Review*, 84: 1-64.
- Searle, J. R. 1990. "Consciousness, explanatory inversion and cognitive science." *Behavioral and Brain Sciences*, 13: 585-642.
- Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Siewert, C. 1998. *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.
- Siewert, C. 2011. "Consciousness and Intentionality", *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2011/entries/consciousness/intentionality/>>.
- Shallice, T. 1988. *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.
- Shear, J. 1997. *Explaining Consciousness: The Hard Problem*. Cambridge, MA: MIT Press.
- Shoemaker, S. 1975. "Functionalism and qualia," *Philosophical Studies*, 27: 291-315.
- Shoemaker, S. 1981. "Absent qualia are impossible." *Philosophical Review*, 90: 581-99.
- Shoemaker, S. 1982. "The inverted spectrum." *Journal of Philosophy*, 79: 357-381.
- Shoemaker, S. 1990. "Qualities and qualia: what's in the mind," *Philosophy and Phenomenological Research*, Supplement, 50: 109-131.
- Shoemaker, S. 1998. "Two cheers for representationalism," *Philosophy and Phenomenological Research*.
- Silberstein, M. 1998. "Emergence and the mind-body problem." *Journal of Consciousness Studies*, 5: 464-82.
- Silberstein, M 2001. "Converging on emergence: consciousness, causation and explanation." *Journal of Consciousness Studies*, 8: 61-98.
- Singer, P. 1975. *Animal Liberation*. New York: Avon Books.
- Singer, W. 1999. "Neuronal synchrony: a versatile code for the definition of relations." *Neuron*, 24: 49-65.
- Skinner, B. F. 1953. *Science and Human Behavior*. New York: MacMillan.

- Smart, J. 1959. "Sensations and brain processes." *Philosophical Review*, 68: 141-56.
- Stapp, H. 1993. *Mind, Matter and Quantum Mechanics*. Berlin: Springer Verlag.
- Stoljar, D. 2001. "Two conceptions of the physical." *Philosophy and Phenomenological Research*, 62: 253-81
- Strawson, G. 1994. *Mental Reality*. Cambridge, Mass: MIT Press, Bradford Books.
- Swinburne, R. 1986. *The Evolution of the Soul*. Oxford: Oxford University Press.
- Titchener, E. 1901. *An Outline of Psychology*. New York: Macmillan.
- Travis, C. 2004. "The silence of the senses." *Mind*, 113: 57-94.
- Triesman, A. and Gelade, G. 1980. "A feature integration theory of attention." *Cognitive Psychology*, 12: 97-136.
- Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
- Tye, M. 2000. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.
- Tye, M. 2003. "Blurry images, double vision and other oddities: new troubles for representationalism?" In A. Jokic and Q. Smith eds., *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.
- Van Gulick, R. 1985. "Physicalism and the subjectivity of the mental." *Philosophical Topics*, 13: 51-70.
- Van Gulick, R. 1992. "Nonreductive materialism and intertheoretical constraint." In A. Beckermann, H. Flohr, J. Kim, eds. *Emergence and Reduction*. Berlin and New York: De Gruyter, 157-179.
- Van Gulick, R. 1993. "Understanding the phenomenal mind: Are we all just armadillos?" In M. Davies and G. Humphreys, eds., *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.
- Van Gulick, R. 1994. "Dennett, drafts and phenomenal realism." *Philosophical Topics*, 22/1-2: 443-56.
- Van Gulick, R. 1995. "What would count as explaining consciousness?" In T. Metzinger, ed. *Conscious Experience*. Paderborn: Ferdinand Schöningh.
- Van Gulick, R. 2000. "Inward and upward: reflection, introspection and selfawareness." *Philosophical Topics*, 28: 275-305.
- Van Gulick, R. 2003. "Maps, gaps and traps." In A. Jokic and Q. Smith eds. *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.

- Van Gulick, R. 2004. "Higher-order global states HOGS: an alternative higher-order model of consciousness." In Gennaro, R. ed. *Higher-Order Theories of Consciousness*. Amsterdam and Philadelphia: John Benjamins.
- Van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- Varela, F. and Maturana, H. 1980. *Cognition and Autopoiesis*. Dordrecht: D. Reidel.
- Varela, F. 1995. "Neurophenomenology: A methodological remedy for the hard problem." *Journal of Consciousness Studies*, 3: 330-49.
- Varela, F. and Thomson, E. 2003. "Neural synchronicity and the unity of mind: a neurophenomenological perspective." In Cleermans, A. ed. *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford: Oxford University Press.
- Velmans, M. 1991. "Is Human information processing conscious?" *Behavioral and Brain Sciences*, 14/4: 651-668
- Velmans, M. 2003. "How could conscious experiences affect brains?" *Journal of Consciousness Studies*, 9: 3-29.
- von Helmholtz, H. 1897/1924. *Treatise on Physiological Optics*. Translated by J. Southly. New York: Optical Society of America.
- Wilkes, K. V. 1984. "Is consciousness important?" *British Journal for the Philosophy of Science*, 35: 223-43.
- Wilkes, K. V. 1988. "Yishi, duo, us and consciousness." In A. Marcel and E. Bisiach, eds., *Consciousness in Contemporary Science*. Oxford: Oxford University Press.
- Wilkes, K. V. 1995. "Losing consciousness." In T. Metzinger, ed. *Conscious Experience*. Paderborn: Ferdinand Schöningh.
- Watson, J. 1924. *Behaviorism*. New York: W. W. Norton.
- Wegner, D. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Wittgenstein, L. 1921/1961. *Tractatus Logico-Philosophicus*. Translated by D. Pears and B. McGuinness. London: Routledge and Kegan Paul.
- Wundt, W. 1897. *Outlines of Psychology*. Leipzig: W. Engleman.
- Yablo, S. 1998. "Concepts and consciousness." *Philosophy and Phenomenological Research*, 59: 455-63.

Sobre Tradutores e Revisores

Verbetes (I) - Condições Necessárias e Suficientes:

Luiz Helvécio Marques Segundo (tradutor): Doutor em filosofia pelo PPG-Fil/UFSC. Atualmente é pesquisador vinculado à Associação Brasileira de Filosofia da Religião (ABFR). Tem atuado na interseção entre epistemologia, filosofia da religião e filosofia da ciência, investigando questões relativas ao debate entre ciência e religião, com particular ênfase à teoria da evolução biológica. Interessa-se também por abordagens evolutivas à moralidade e à cultura.

Sérgio Ricardo Neves de Miranda (revisor): Mestre em Filosofia pela Universidade Federal de Minas Gerais (1997) e doutorado em Filosofia pela Universidade de Bielefeld (2006). Atualmente é professor associado do Departamento de Filosofia na Universidade Federal de Ouro Preto. Áreas de interesse: Epistemologia, Metafísica e Filosofia da Religião.

Verbetes (II) - Referência:

Lucas Roisenberg Rodrigues (tradutor): Graduado em Filosofia pela Universidade Federal do Rio Grande do Sul (UFRGS). Possui mestrado em Filosofia pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Possui doutorado em Filosofia Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) Em seu mestrado tratou do paradoxo do prefácio, e de temas relacionados como o papel da lógica na determinação da racionalidade e da justificação. Em seu doutorado, fez uma análise crítica das epistemologias modais. Trabalhou como professor de filosofia na Universidade Federal Da Fronteira SUL (UFFS). Principais áreas de interesse: metafísica, epistemologia, lógica e filosofia da linguagem.

Jerzy André Brzozowski (revisor): Mestre e doutor em Filosofia, ambos pela Universidade Federal de Santa Catarina. Sua tese de doutorado diz respeito à natureza dos nomes próprios teóricos na biologia evolutiva. Entre 2011 e 2018, foi professor adjunto na Universidade Federal da Fronteira Sul -- Campus Erechim. Em 2018, foi pesquisador visitante na University of Utah. Atualmente, é

professor adjunto do Departamento de Filosofia na Universidade Federal de Santa Catarina. Atua sobretudo na área de Filosofia, particularmente em Teoria do Conhecimento, Filosofia da Ciência, Filosofia da Linguagem e Filosofia da Biologia. Tem pesquisado alguns temas da relação entre conhecimento científico e verdades modais.

Verbetes (III) - A Metafísica da Causação:

Rodrigo Reis Lastra Cid (tradutor): Professor Adjunto de Filosofia na Universidade Federal do Amapá. Pesquisador na área da Metafísica. Editor Chefe do periódico *Investigação Filosófica*, Editor da Série *Investigação Filosófica* (NEP-FIL/Editora UFPel). Líder do Grupo de Pesquisa *Investigação Filosófica* (DGP/CNPq), membro do GT de Metafísica Analítica da ANPOF, da *Society for the Metaphysics of Science* e do Laboratório de Estudos Aplicados em Infoprodutos (UFES/CNPq). Residência Pós-Doutoral em Filosofia realizada na Universidade Federal de Minas Gerais. Doutor e Mestre em Lógica e Metafísica (Filosofia) pelo Programa de Pós-Graduação em Lógica e Metafísica da Universidade Federal do Rio de Janeiro, com período sanduíche na *Université Catholique de Louvain*. Bacharel em Filosofia pela Universidade Federal de Ouro Preto. Ex-Professor de Filosofia na Universidade Federal do Rio de Janeiro, no Instituto Federal de Minas Gerais e na Faculdade Dom Luciano Mendes.

Bruno Borges Moura (tradutor): Graduando em Filosofia pela UNIFAP, iniciação científica em tradução e membro do Grupo de Pesquisa *Investigação Filosófica* e secretário do Projeto de Extensão *Jovem Bilíngue* na UNIFAP.

Pedro Merluzzi (revisor): Doutor em filosofia pela Universidade de Durham (orientadora: Nancy Cartwright; examinadores: Helen Beebee & Matthew Tugby). Tem interesse nas áreas de metafísica, filosofia da ciência, lógica filosófica e história da filosofia moderna. Atualmente é bolsista de pós-doutorado (FAPESP) pelo Centro de Lógica, Epistemologia e História da Ciência da UNICAMP sob a supervisão de Marco Ruffino.

Verbetes (IV) - Consciência:

Marcelo Fischborn (tradutor): Professor de Filosofia no Instituto Federal Farroupilha, Campus Avançado Uruguaiana. Realizei pesquisa pós-doutoral em

filosofia (bolsa Capes/Fapergs) na Universidade Federal de Santa Maria (UFSM), instituição onde também realizei doutorado (2018), mestrado (2014) e licenciatura (2011) em filosofia. Durante o doutorado, fui pesquisador visitante na Florida State University através do Programa Estágio de Doutorando nas Ciências Humanas, Ciências Sociais, Letras e Artes nos EUA (financiado pela CAPES e pela Comissão Fulbright Brasil). Minha pesquisa tem como preocupação central a estruturação de uma investigação interdisciplinar sobre o aprimoramento das práticas de responsabilização. Áreas e tópicos de interesse incluem: ética, psicologia moral, filosofia da mente e da ação, livre-arbítrio, responsabilidade moral, censura, punição e violência.

Marco Aurelio Sousa Alves (tradutor): Professor Adjunto (DFIME / UFSJ); Professor Permanente (PPGFIL / UFSJ); Professor Permanente (POSDEFIL / UFOP) Lecionou anteriormente nos departamentos de filosofia da UFMG, FAJEMG, PUC-MG, e University of Texas at Austin (EUA). Ph.D. em Filosofia pela University of Texas at Austin (EUA). Mestrado e Graduação em Filosofia pela UFMG. Trabalha com filosofia da mente e da percepção, filosofia da linguagem, epistemologia, metafísica, filosofia da ciência, lógica e ética.

André Joffily Abath (revisor): Professor adjunto IV no Departamento de Filosofia da Universidade Federal de Minas Gerais. Possui doutorado em Filosofia pela Universidade de Sheffield (2007), mestrado em Filosofia da Linguagem e da Consciência pela Universidade de Lisboa (2003), graduação em Filosofia pela Universidade Federal da Paraíba (2000), e graduação em Comunicação Social pela Universidade Federal da Paraíba (1999). Pesquisa nas áreas de epistemologia, filosofia da mente e filosofia da linguagem. Foi diretor da Revista *Kriterion* (2014-2016) e tutor do Programa de Educação Tutorial (PET) em Filosofia da UFMG (2016-2018).



Editora
UFPeI

DISSERTATIO
FILOSOFIA