Editora
UFPel

Série Investigação Filosófica

Textos selecionados de

Metafísica

Pedro Merlussi
(Organizador)

DISSERTATIO
FILOSOFIA

TEXTOS SELECIONADOS DE METAFÍSICA

Série Investigação Filosófica

TEXTOS SELECIONADOS DE METAFÍSICA

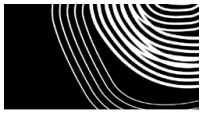
Pedro Merlussi
(Organizador)

Pelotas, 2021

REITORIA
Reitora: Isabela Fernandes Andrade
Vice-Reitora: Ursula Rosa da Silva
Chefe de Gabinete: Aline Ribeiro Paliga
Pró-Reitor de Graduação: Maria de Fátima Cóssio
Pró-Reitor de Pesquisa e Pós-Graduação: Flávio Fernando Demarco
Pró-Reitor de Extensão e Cultura: Eraldo dos Santos Pinheiro
Pró-Reitor de Planejamento e Desenvolvimento: Paulo Roberto Ferreira Júnior
Pró-Reitor Administrativo: Ricardo Hartlebem Peter
Pró-Reitor de Gestão de Informação e Comunicação: Julio Carlos Balzano de Mattos
Pró-Reitor de Assuntos Estudantis: Fabiane Tejada da Silveira
Pró-Reitor de Gestão Pessoas: Taís Ulrich Fonseca

CONSELHO EDITORIAL DA EDITORA DA UFPEL
Presidente do Conselho Editorial: Ana da Rosa Bandeira
Representantes das Ciências Agronômicas: Victor Fernando Büttow Roll
Representantes da Área das Ciências Exatas e da Terra: Eder João Lenardão
Representantes da Área das Ciências Biológicas: Rosângela Ferreira Rodrigues
Representante da Área das Engenharias e Computação: Reginaldo da Nóbrega Tavares
Representantes da Área das Ciências da Saúde: Fernanda Capella Rugno
Representante da Área das Ciências Sociais Aplicadas: Daniel Lena Marchiori Neto
Representante da Área das Ciências Humanas: Charles Pereira Pennaforte
Representantes da Área das Linguagens e Artes: Lúcia Bergamaschi Costa Weymar

EDITORA DA UFPEL
Chefia: Ana da Rosa Bandeira (Editora-chefe)
Seção de Pré-produção: Isabel Cochrane (Administrativo)
Seção de Produção: Suelen Aires Böettge (Administrativo)
Anelise Heidrich (Revisão)
Ingrid Fabiola Gonçalves (Diagramação)
Seção de Pós-produção: Madelon Schimmelpfennig Lopes (Administrativo)
Morgana Riva (Assessoria)



CONSELHO EDITORIAL
Prof. Dr. João Hobuss (Editor-Chefe)
Prof. Dr. Juliano Santos do Carmo (Editor-Chefe)
Prof. Dr. Alexandre Meyer Luz (UFSC)
Prof. Dr. Rogério Saucedo (UFSM)
Prof. Dr. Renato Duarte Fonseca (UFSM)
Prof. Dr. Arturo Fatturi (UFFS)
Prof. Dr. Jonadas Techio (UFRGS)
Profa. Dra. Sofia Albornoz Stein (UNISINOS)
Prof. Dr. Alfredo Santiago Culleton (UNISINOS)
Prof. Dr. Roberto Hofmeister Pich (PUCRS)
Prof. Dr. Manoel Vasconcellos (UFPEL)
Prof. Dr. Marco Antônio Caron Ruffino (UNICAMP)
Prof. Dr. Evandro Barbosa (UFPEL)
Prof. Dr. Ramón del Castillo (UNED/Espanha)
Prof. Dr. Ricardo Navia (UDELAR/Uruguai)
Profa. Dra. Mónica Herrera Noguera (UDELAR/Uruguai)
Profa. Dra. Mirian Donat (UEL)
Prof. Dr. Giuseppe Lorini (UNICA/Itália)
Prof. Dr. Massimo Dell'Utri (UNISA/Itália)

COMISSÃO TÉCNICA (EDITORAÇÃO)
Prof. Dr. Juliano Santos do Carmo (Diagramador/Capista)

DIREÇÃO DO IFISP
Prof. Dr. João Hobuss

CHEFE DO DEPARTAMENTO DE FILOSOFIA
Prof. Dr. Juliano Santos do Carmo

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 (IF/UNIFAP)
Juliano Santos do Carmo (NEPFIL/UFPEL)

COMISSÃO TÉCNICA
Marco Aurélio Scarpino Rodrigues (Revisor em Língua Portuguesa)
Rafaela Nobrega (Diagramadora/Capista)

ORGANIZADOR DO VOLUME
Pedro Merlussi (UNICAMP)

TRADUTORES E REVISORES
Alex Bispo (UNICAMP)
Ariston Sócrates G. S. Cruz (UNICAMP)
Gabriel A. Eugênio (UNICAMP)
Karenina do Nascimento Rodrigues (UNICAMP)
Kherian Gracher (UFSC)
Renato Semaniuc Valvassori (UNICAMP)
Rhamon de Oliveira Nunes (UERN)
Ulisses Franceschi Eliano (UNICAMP)
Vinicius Bianchi (UFSC)

CRÉDITO DA IMAGEM DE CAPA
Jacques Louis David. 1784. Source: Wikimedia Commons

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

Aluízio 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 Cecílio
Kherian Galvão Cesar Gracher
Luiz Helvécio Marques Segundo
Paulo Roberto Moraes de Mendonça
Pedro Merlussi
Rafael César Pitt
Rafael Martins
Renata Ramos da Silva
Rodrigo Alexandre de Figueiredo
Rodrigo Reis Lastra Cid
Sagid Salles
Tiago Luís Teixeira de Oliveira

© Série Investigação Filosófica, 2021

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

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

Os direitos autorais estão de acordo com a Política Editorial do NEPFil 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 2021 por NEPFil online e Editora da UFPel.

Dados Internacionais de Catalogação

N123 Textos selecionados de metafísica.
[recurso eletrônico] Organizador: Pedro Merlussi – Pelotas: NEPFIL Online, 2021.

259p. - (Série Investigação Filosófica).

Modo de acesso: Internet
<wp.ufpel.edu.br/nepfil>
ISBN: 978-65-86440-35-5

1. Filosofia. 2. Metafísica I. Merlussi, Pedro.

COD 100

creative commons br
BY NC ND

ABEC BRASIL
Associação Brasileira de Editores Científicos

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

SUMÁRIO

Sobre a série Investigação Filosófica **13**
Introdução **15**

(I) Argumentos a favor do Incompatibilismo **17**
1. Considerações Preliminares 19
2. Duas Razões para Pensar que o Livre-Arbítrio é Incompatível com o Determinismo 27
3. Argumentos Baseados em Intuição 36
3.1 O Argumento da Ausência de Bifurcações 36
3.2 Argumentos da Manipulação 38
4. Argumentos da Origem 44
5. Escolha e o Argumento da Consequência 49
Referências 60

(II) Responsabilidade Moral **79**
1. Liberdade, Responsabilidade e Determinismo 82
2. Algumas Teorias sobre a Responsabilidade Moral 87
2.1 Teorias Prospectivas 87
2.2 A Teoria das Atitudes Reativas 90
2.2.1 "Liberdade e Ressentimento" 90
2.2.2 As Críticas à Teoria de Strawson 93
2.3 As Teorias da Sensibilidade às Razões 95
3. Debates Contemporâneos 98
3.1 As "Faces" da Responsabilidade 98
3.1.1 Atributabilidade *versus* Imputabilidade 98
3.1.2 Atribuicionismo 101
3.1.3 Responsividade 103
3.2 Competência Moral 103
3.2.1 A Condição da Competência Moral sobre Responsabilidade 103
3.2.2. Teorias Conversacionais sobre a Responsabilidade 105
3.2.3 Psicopatia 106
3.3 Ceticismo e Tópicos Relacionados 107
3.3.1 Sorte Moral 108

3.3.2 A Responsabilidade Última 110
3.3.3 História Pessoal e Manipulação 113
3.3.4 A Condição Epistêmica sobre Responsabilidade 117
Referências 121

(III) Compatibilismo 136
1. Livre-Arbítrio e o Problema do Determinismo Causal 137
1.1 Determinismo e Possibilidades Alternativas 139
1.2 Determinismo e Origem 140
1.3 Respostas Compatibilistas 142
2. Compatibilismo Clássico 143
2.1 Liberdade de acordo com o Compatibilismo Clássico 143
2.2 A Análise Condicional Compatibilista Clássica 144
2.3 A Última Influência da Análise Condicional 146
3. Compatibilismo em Transição 147
3.1 O Argumento da Consequência 147
3.2 Um Desafio ao Princípio de Possibilidades Alternativas 149
3.3 O Foco nas Atitudes Reativas 151
4. Compatibilismo Contemporâneo 152
4.1 Compatibilismo em relação à Liberdade para Agir de Outro Modo 152
4.1.1 Desafiando o Controle Necessário e o Passado 153
4.1.2 Desafiando o Controle Necessário e as Leis da Natureza 154
4.1.3 Desafiando as Inferências baseadas no Controle Necessário 155
4.1.4 Dando uma Explicação sobre a Liberdade de Agir de Outro Modo 156
4.1.5 O Novo Disposicionalismo 156
4.2 Compatibilismo Hierárquico 158
4.2.1 Desejos de Ordem-Superior e Livre-Arbítrio 159
4.2.2 Dois Problemas para a Teoria Hierárquica 161
4.3 A Perspectiva da Razão 162
4.4 Compatibilismo de Responsividade a Razões 164
4.4.1 Responsividade a Razões baseada no Agente 165
4.4.2 Uma Tensão entre Capacidade de Resposta a Razões e os Exemplos de Frankfurt 165
4.4.3 Uma Teoria de Responsividade a Razões baseada em Mecanismos 166

4.4.4 Avaliando o Compatibilismo de Responsividade a Razões 168
4.5 Compatibilismo Strawsoniano 169
Referências 170

(IV) Teorias Contrafactuais de Causação 181
1. A Análise de Lewis para Contrafactuais de 1973 182
1.1 Contrafactuais e Dependência Causal 183
1.2 A Assimetria Temporal da Dependência Causal 185
1.3 Transitividade e Preempção 188
1.4 Causação Incerta 189
2. Problemas para a Teoria Contrafactual de Lewis 191
2.1 Assimetria Temporal 191
2.2 Transitividade 193
2.3 Preempção 195
3. A Teoria de Lewis de 2000 197
4. Contextualismo *versus* Invariantismo 200
5. A Modelagem de Equações Estruturais 204
5.1 MEE: Introdução Básica 204
5.2 MEE e Contrafactuais 212
5.3 Modelos e Realidade 213
5.4 MEE e Causação Incerta 215
5.5 Padrões e Desvios 216
Referências 219

(V) As Leis da Natureza 226
1. A Questão Básica: O que é uma Lei? 227
2. Os Sistemas 229
3. Universais 231
4. A Sobreveniência Humiana 233
5. Antirrealismo 236
6. Antirreducionismo 238
7. Indução 239
8. Necessidade 242
9. Leis, Circularidade e as Perspectivas de Explicação 243
10. Física e as Ciências Especiais 246
10.1. A Física tenta descobrir Regularidades sem Exceções? 246

10.2. Poderia haver alguma Lei de Ciência Especial? 248
11. Considerações Finais: o que vem em seguida? 250
Referências 251

Sobre os tradutores e revisores 255

SOBRE A SÉRIE INVESTIGAÇÃO FILOSÓFICA

A *Série Investigação Filosófica* é uma coleção de livros de traduções de verbetes da *Enciclopédia de Filosofia de Stanford* (*Stanford Encyclopedia of Philosophy*), que se intenciona a servir tanto como material didático, para os professores das diferentes subáreas 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 bons materiais em português para indicarmos aos estudantes, e há uma certa deficiência na graduação brasileira de Filosofia, principalmente em localizações menos favorecidas, em relação ao conhecimento de outras línguas, como o inglês e o francês. Sendo assim, tentamos suprir essa deficiência, introduzindo essas traduções ao público de Língua Portuguesa, sem nenhuma finalidade comercial, meramente pela glória da Filosofia. Aproveitamos para agradecer a John Templeton Foundation por financiar a publicação de vários dos livros de nossa série, incluindo este, e eximi-la de quaisquer opiniões aqui contidas, as quais são de responsabilidade de seus devidos autores. [*This publication was made possibile through a support of a grant from John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.*]

Essas traduções foram todas realizadas por filósofos ou por estudantes de filosofia supervisionados, além de, posteriormente, terem sido revisadas por especialistas nas respectivas áreas. Todas as traduções dos verbetes foram autorizadas pelo querido Prof. Dr. Edward Zalta, editor da *Enciclopédia de Filosofia de Stanford*, razão pela qual o agradecemos imensamente. Sua disposição em contribuir para a ciência brinda os países de Língua Portuguesa com um material filosófico de excelência, disponibilizado gratuitamente no site da Editora da Universidade Federal de Pelotas (UFPel), assim, contribuindo para nosso maior princípio, a ideia de transmissão de conhecimento livre, além de, também, corroborar nossa intenção, a de promover o desenvolvimento da Filosofia em Língua Portuguesa e seu ensino no país. Aproveitamos o ensejo para agradecer, também, ao editor da UFPel, na figura do Prof. Dr. Juliano do Carmo, que apoiou nosso projeto desde o

início. Agradecemos, ainda, a todos os organizadores, tradutores e revisores, que participam de nosso projeto. Sem a dedicação voluntária desses colaboradores, nosso trabalho não teria sido possível. Esperamos, com o início desta Série, abrir as portas para o crescimento desse projeto de tradução e trabalharmos em conjunto pelo crescimento da Filosofia em Língua Portuguesa.

Prof. Dr. Rodrigo Reis Lastra Cid (IF/UNIFAP)
Prof. Dr. Juliano Santos do Carmo (NEFIL/UFPEL)
Editores da *Série Investigação Filosófica*

INTRODUÇÃO

Os verbetes selecionados para este volume lidam com alguns dos problemas centrais da metafísica na filosofia analítica contemporânea: O que é o livre-arbítrio? Será que o livre-arbítrio é compatível com o determinismo? Uma pessoa precisa ter livre-arbítrio para ser moralmente responsável? O que é a responsabilidade moral? O que significa dizer que um evento causou outro? Podemos explicar a causalidade por meio de condicionais contrafactuais? O que é uma lei da natureza? As leis da natureza governam os acontecimentos do nosso mundo?

A ideia de organizar o volume com as presentes traduções resultou de um curso que ministrei na Universidade Estadual de Campinas sobre a metafísica do livre-arbítrio em 2019. Os três primeiros verbetes servem para suplementar o material disponível em português sobre os conceitos de livre-arbítrio e responsabilidade moral. Não obstante, acrescentei a este volume os verbetes sobre as leis da natureza e as teorias contrafactuais da causação, pois os problemas da metafísica das leis da natureza e da causalidade afetam diretamente os debates sobre o livre-arbítrio. Em todo o caso, os verbetes selecionados para este volume devem também servir para suplementar o material disponível em português de tópicos mais gerais, como condicionais contrafactuais, causalidade, leis da natureza, etc.

O debate entre compatibilistas e incompatibilistas é um debate sobre a questão de saber se o livre-arbítrio é compatível com o determinismo. Esta questão é discutida em detalhes nos verbetes "Argumentos a favor do Incompatibilismo", de Kadri Vihvelin, e "Compatibilismo", de Michael McKenna e Justin Coates. O principal argumento incompatibilista na literatura atual é o Argumento da Consequência, discutido nos verbetes mencionados. Este argumento repousa em uma premissa sobre as leis da natureza, a saber, a premissa de que *não temos o poder* de mudar as leis da natureza, a qual foi questionada na literatura de maneira particularmente influente por David Lewis. No entanto, há uma importante conexão entre este debate sobre o conceito de livre-arbítrio e o debate sobre a metafísica das leis. Será que diferentes teorias das leis da natureza afetam os debates sobre o conceito de livre-

arbítrio? Neste sentido, o último verbete deste volume, “Leis da Natureza”, de John Carroll, visa preencher esta lacuna sobre as possíveis conexões acerca dos debates do livre-arbítrio e da natureza das leis.

O verbete “Teorias Contrafactuais da Causação” de Peter Menzies e Helen Beebee contribui também ao verbete já publicado verbete sobre causação. O debate sobre causação é particularmente relevante nas discussões sobre o livre-arbítrio e a responsabilidade moral, sobretudo após Harry Frankfurt influentemente argumentar que uma pessoa pode ser moralmente responsável por uma ação mesmo se for incapaz de agir de outro modo. O verbete “Responsabilidade Moral”, de Matthew Talbert, suplementa esta discussão e mostra como os debates sobre a responsabilidade moral se descolaram dos debates sobre a natureza do livre-arbítrio, sobretudo após as influentes abordagens de Frankfurt e Strawson.

Em suma, embora o pano de fundo dos verbetes seja a metafísica do livre-arbítrio, os verbetes desta edição também preenchem lacunas na bibliografia disponível em português de alguns dos principais problemas da metafísica contemporânea na filosofia analítica. Diversas pessoas trabalharam neste projeto e sou extremamente grato a todas elas: Alex Bispo, Ariston Sócrates G. S. Cruz, Gabriel A. Eugênio, Gabriel Maruchi, Karenina do Nascimento Rodrigues, Kherian Gracher, Renato Semaniuc Valvassori, Rhamon de Oliveira Nunes e Ulisses Franceschi Eliano. Muito obrigado!

Pedro Merlussi

(I) Argumentos a favor do Incompatibilismo*

Autor: Kadri Vihvelin
Tradutores: Gabriel de Andrade Pagnozzi Maruchi & Pedro Merlussi
Revisores Científicos: Kherian Gracher & Rhamon de Oliveira Nunes

Acreditamos que temos livre-arbítrio e esta crença está tão fortemente enraizada em nosso cotidiano que é quase impossível levar a sério a possibilidade de que esteja errada. Por exemplo, no processo de deliberação, assumimos que há mais de uma escolha que possamos fazer, mais de uma ação que somos capazes de executar. Quando olhamos para o passado e nos arrependemos de uma decisão ingênua, ou nos culpamos por não termos feito algo que deveríamos, assumimos que poderíamos ter escolhido e agido de maneira diferente da qual agimos. Quando olhamos para o futuro e fazemos planos, assumimos que temos ao menos algum controle sobre nossas ações e o curso de nossas vidas, pensamos que, ao menos

---

* VIHVELIN, K. Arguments for Incompatibilism. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Fall Edition. Stanford, CA: The Metaphysics Research Lab, 2018. Disponível em: https://plato.stanford.edu/archives/fall2018/entries/incompatibilism-arguments/. Acesso em: 01 ago. 2021.

The following is the translation of the entry on Scientific Method, 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/fall2018/entries/incompatibilism-arguments/>. 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//incompatibilism-arguments/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry.

às vezes, o que escolhemos e tentamos fazer está em nossas mãos.

O determinismo é uma tese bem geral sobre o universo: simplificando bastante, é a tese de que tudo o que acontece, incluindo tudo o que escolhemos e fazemos, é determinado por fatos sobre o passado em conjunção com as leis da natureza. O determinismo não pertence ao bom senso, e, dado tudo o que sabemos, não é fácil levar a sério a possibilidade de que possa ser verdadeiro. A(o) incompatibilista acredita que, se no final das contas, o determinismo for verdadeiro, então nossa crença de que temos livre-arbítrio é falsa. A(o) compatibilista nega que a verdade do determinismo tenha essa consequência drástica. Nesta perspectiva, a verdade do determinismo é compatível com a verdade da nossa crença de que temos livre-arbítrio. O problema filosófico do livre-arbítrio e do determinismo é o problema de decidir quem está certo: a(o) compatibilista ou a(o) incompatibilista.

Boa parte do interesse filosófico no problema do livre-arbítrio/determinismo se motiva através de considerações sobre a responsabilidade moral, já que é geralmente aceito que a posse do livre-arbítrio seja uma condição necessária para a responsabilidade moral. Logo, se o determinismo impedisse a existência do livre-arbítrio, também impediria a responsabilidade moral. Porém, é importante distinguir as questões sobre o livre-arbítrio (se o temos, o que é, se é compatível com o determinismo, se é compatível com outras coisas que acreditamos ser verdadeiras) das questões sobre a responsabilidade moral. Alguém pode acreditar que temos livre-arbítrio e que o livre-arbítrio é compatível com o determinismo ao mesmo tempo em que acredita que, por outros motivos, ninguém é moralmente responsável. E alguém pode acreditar que não temos livre-arbítrio, por causa do determinismo ou outra tese, enquanto também acredita, em oposição ao bom senso, que, mesmo assim, somos moralmente responsáveis. O que uma pessoa acredita sobre o determinismo e a responsabilidade moral dependerá, em grande parte, do que essa pessoa acredita sobre os vários assuntos que estão no escopo da ética em vez da metafísica (P. STRAWSON, 1962; G. STRAWSON, 1986, 1994; WATSON, 2004). Por esses e outros motivos, é importante não confundirmos a questão da compatibilidade entre o livre-arbítrio e o determinismo com a questão de saber se a responsabilidade moral é compatível com o determinismo.

É preciso reconhecer que a literatura mudou as definições, e muitos teóricos e teóricas contemporâneos entendem os termos “compatibilismo” e “incompatibilismo” como teses sobre a responsabilidade moral (ou “liberdade moral” ou a liberdade que “fundamenta” ou “explica” a responsabilidade moral) ao invés de teses sobre o livre-arbítrio (PEREBOOM, 1995, 2001, 2014). Confira Vihvelin (2011) para uma

discussão sobre o tema. Este verbete não seguirá esse uso.

Este verbete é sobre o problema do livre-arbítrio/determinismo, mais especificamente é sobre os argumentos pela tese de que o livre-arbítrio é incompatível com o determinismo. Dado que a literatura contemporânea sobre o livre-arbítrio está em grande parte dominada por considerações sobre a responsabilidade moral, alguns dos argumentos considerados serão pela tese de que, se o determinismo for verdadeiro, nunca somos moralmente responsáveis porque nunca satisfazemos a "condição de liberdade" para sermos moralmente responsáveis por nossas ações. Não obstante, o foco deste verbete será saber se o livre-arbítrio, ou agir com livre-arbítrio, é compatível com o determinismo.

1. Considerações Preliminares

Na literatura, "determinismo" é amiúde utilizado como um termo guarda-chuva para uma variedade de teses que, tradicionalmente, são consideradas como ameaças ao livre-arbítrio. Dado esse uso, a tese que estamos chamando de "determinismo" (determinismo nomológico, também denominado "determinismo causal") é apenas um dos diferentes tipos de determinismos, e o problema do livre-arbítrio/determinismo que discutiremos pertence a uma família de problemas relacionados. Por exemplo, o determinismo lógico é a tese de que o princípio de bivalência vale para todas as proposições, incluindo as proposições sobre o futuro, e o problema do livre-arbítrio e do determinismo lógico é o problema de decidir se nossa crença de que temos livre-arbítrio é compatível com a existência de verdades sobre nossas ações futuras. (Deve-se reconhecer que, neste ponto, há um consenso quase universal entre filósofos e filósofas de que a resposta é "sim"). Para mais informações sobre o determinismo lógico, conferir a entrada sobre o **fatalismo**, além dos trabalhos de Taylor, de 1962 e 1963 (1992), D. Lewis (1976), Westphal (2006), Merricks (2009), Fischer e Todd (2011), Vihvelin (2013). O determinismo teológico é a tese de que Deus existe e tem conhecimento infalível de todas as proposições verdadeiras, incluindo as proposições sobre nossas ações futuras. O problema do livre-arbítrio e do determinismo teológico é entender como, e se é possível, existir o livre-arbítrio se Deus (que não pode estar enganado) sabe o que iremos fazer. Para mais informações sobre o determinismo teológico, conferir os verbetes sobre o **fatalismo** e a **pré-ciência divina**, e Fischer e Todd (2015).

Neste verbete, restringiremos nossa atenção a argumentos pela incompatibilidade entre o livre-arbítrio e o determinismo nomológico, mas é importante entendermos um ponto preliminar. O determinismo lógico e o determinismo nomológico são duas teses bastante diferentes. A primeira diz respeito à **verdade**, a última diz respeito **às leis da natureza**. O determinismo lógico nada diz sobre a causação ou as leis da natureza, simplesmente diz que, independentemente de sabermos quais são as leis da natureza, há proposições atemporalmente verdadeiras sobre o passado, presente e futuro, incluindo nossas ações futuras. O determinismo lógico não implica o determinismo nomológico, pois o primeiro pode ser verdadeiro mesmo que o último seja falso. Por sua vez, o determinismo nomológico diz, grosso modo, que os fatos sobre o passado em conjunção com os fatos sobre as leis da natureza determinam todos os fatos sobre o futuro. Logo, se o determinismo nomológico for verdadeiro, há proposições verdadeiras sobre nossas ações futuras.

Por que isso importa? Pela seguinte razão. Uma típica resposta inicial ao determinismo lógico é a resposta fatalista:

> Se há verdades sobre o que farei no futuro, então **devo** fazer o que quer que seja que **irei** fazer, portanto não tenho livre-arbítrio.

Ainda que esta resposta tenha um forte apelo intuitivo (como pode ser visto em grande parte das discussões de cultura popular, e mesmo científicas, sobre viagem no tempo), em geral, é aceito na Filosofia que o fatalista está cometendo um erro. Diferentes diagnósticos foram dados da "falácia fatalista", ou outro erro no raciocínio modal, mas o ponto básico é simples: verdade não é o mesmo que necessidade; não é o mesmo que necessidade lógica, necessidade metafísica, ou mesmo a necessidade relativa da inevitabilidade ou falta de capacidade ou poder. A existência de um conjunto detalhado de verdades sobre minhas ações futuras é consistente com a minha capacidade de fazer coisas diferentes das coisas que realmente fiz.

É claro que é possível aceitar que a existência de **verdades** sobre nossas ações futuras é compatível com o livre-arbítrio e, ao mesmo tempo, negar que a existência de verdades **nomologicamente determinadas** sobre nossas ações futuras seja compatível com o livre-arbítrio. Entretanto, precisamos de um argumento a favor dessa conclusão, um argumento que não dependa do raciocínio fatalista ou que apele a intuições fatalistas.

Para comparações entre os argumentos a favor do incompatibilismo e os argumentos a favor do fatalismo, consulte van Inwagen (1983), Mackie (2003), Perry (2004) e Vihvelin (2008, 2013).

À primeira vista, o determinismo nomológico, doravante "determinismo", é uma tese contingente e empírica sobre as leis da natureza: que elas são determinísticas em vez de probabilísticas, e que cobrem todos os eventos em vez de terem um escopo limitado. Numa observação mais atenta, as leis são determinísticas se elas implicam regularidades sem exceções (por exemplo, que todos os **F**s são **G**s, que todos os **ABCD**s são **E**s) e as leis são probabilísticas se dizem que **F**s tem uma probabilidade objetiva **N** (menor que 1) de serem **G**s. Observe que as ditas "leis estatísticas" não precisam ser leis probabilísticas (*vide* ARMSTRONG, 1983, LOEWER, 1996a). As leis da natureza cobrem tudo se as leis determinísticas ou probabilísticas se aplicam a tudo no universo, sem nenhuma exceção. Se, por outro lado, alguns indivíduos ou algumas partes de indivíduos, como as mentes não-físicas dos seres humanos, ou alguns dos comportamentos de alguns indivíduos, tais como as ações livres de seres humanos, não forem cobertos por leis determinísticas ou probabilísticas, então as leis têm escopo limitado em vez de cobrirem tudo.

Para uma articulação mais precisa do determinismo, a literatura contemporânea nos oferece duas opções principais.

Seguindo van Inwagen (1983), podemos entender o determinismo como a tese de que o mundo é governado por um conjunto de leis naturais que é tal que qualquer mundo possível com as mesmas leis que o nosso e que é exatamente como nosso mundo em algum instante temporal é exatamente como nosso mundo em todos os outros instantes. Este jeito de entender o determinismo é usualmente chamado de "a definição por Implicação", porque diz que o determinismo é uma relação de implicação entre, de um lado, proposições verdadeiras sobre leis e proposições verdadeiras sobre fatos particulares em um tempo, e, de outro, proposições verdadeiras sobre fatos particulares. Confira, também, Ginet (1990), Beebee e Mele (2002) e o verbete sobre **determinismo causal**.

Alternativamente, seguindo D. Lewis (1973), podemos entender o determinismo como a tese de que nosso mundo é governado por um conjunto de leis, tal que **quaisquer dois mundos possíveis com as nossas leis** que são exatamente similares em algum instante temporal também são exatamente similares em qualquer outro instante (EARMAN, 1986). Essa segunda definição de determinismo é mais forte do que a primeira: se um mundo é determinístico de acordo com a definição Lewis/Earman, então, ele é determinístico de acordo com a definição por Implicação, mas não vice-versa.

Vamos chamar um mundo possível de “determinístico” se, e somente se, a tese do determinismo é verdadeira nesse mundo, e ‘não-determinístico’ se, e somente se, a tese do determinismo for falsa nesse mundo. Há dois modos bem diferentes em que um mundo pode ser não-determinístico. Um deles pode ser não-determinístico porque pelo menos algumas de suas leis fundamentais são probabilísticas, ou um outro pode ser não-determinístico, porque ele não tem leis ou porque suas leis não cobrem tudo. Vamos chamar mundos que são não-determinísticos apenas do primeiro jeito de “mundos probabilísticos” e vamos chamar mundos, que são não-determinísticos do segundo jeito, “sem leis” ou “parcialmente sem leis”.

O determinismo é uma tese sobre as asserções ou as proposições que são as leis do nosso mundo. Ela não diz nada sobre se essas asserções ou proposições são conhecíveis para seres finitos, ainda menos sobre se podem, em princípio, ser usadas para prever eventos futuros. Para mais informações sobre a relação entre o determinismo e preditibilidade, consulte o verbete sobre o **determinismo causal**.

O determinismo, entendido de acordo com qualquer uma das duas definições anteriores, é neutro com respeito a diferentes teorias sobre leis da natureza, indo desde a chamada teoria “regularista ingênua” (SWARTZ, 1986) até teorias humianas em geral ou teorias “dos melhores sistemas” (D. LEWIS, 1973; EARMAN, 1986; LOEWER, 1996a; BEEBEE, 2000; SCHAFFER, 2008) até vários tipos de teorias necessitaristas (S. SHOEMAKER, 1980; ARMSTRONG, 1983; CARROLL, 2008). No livro, que definiu boa parte do cenário da discussão contemporânea sobre o problema do livre-arbítrio/determinismo, van Inwagen (1983) apelou apenas para dois, relativamente incontroversos, pressupostos sobre as leis da natureza: que as proposições, que são leis, têm este status independentemente da história ou do estado presente do conhecimento científico e que as leis impõem limites às nossas capacidades. A maior parte dos participantes do debate atual concorda, mas há dissidentes que pensam que o determinismo precisa ser definido de um jeito metafisicamente mais forte, como um tipo de relação de necessitação entre eventos particulares (MUMFORD; ANJUM, 2014). Para críticas, confira Mackie (2014a) e Franklin (2014).

O determinismo, entendido segundo qualquer uma das duas definições anteriores, não é uma tese sobre causação, não é a tese de que a causação sempre é uma relação entre eventos, e não é a tese de que todo evento tem uma causa. Se, como muitas pessoas acreditam hoje, as leis fundamentais são probabilísticas ao invés de determinísticas, isso não significa que não há causação, mas apenas que temos que revisar nossas teorias de causação para se adequarem aos fatos.

E é isto que filósofos da causação têm feito. Há versões probabilísticas de teorias da causação de "leis governistas", de teorias contrafactuais de causação, e, assim por diante, para todas as principais teorias de causação. Confira os verbetes sobre **metafísica da causação** e **teorias contrafactuais da causação**. Hoje em dia, é geralmente aceito que pode ser verdade que todo evento tem uma causa mesmo se o determinismo for falso e, consequentemente, alguns eventos não terão causas suficientes ou determinísticas.

Uma controvérsia maior é saber se pode ser verdade que todo evento tem uma causa mesmo se nosso mundo não for nem determinístico nem probabilístico. Se pode haver causas sem leis, ou seja, se um evento particular, objeto, ou pessoa pode ser uma causa, por exemplo, sem instanciar uma lei, então, pode ser verdade, mesmo em um mundo sem lei ou parcialmente sem lei, que todo evento tem uma causa (ANSCOMBE, 1981; VAN INWAGEN, 1983).

É menos claro se o determinismo implica a tese de que todo evento tem uma causa. Se implica ou não, isso depende de qual é a teoria correta da causação, de modo particular, depende do que a teoria correta diz sobre a relação entre causação e leis da natureza.

O que é claro, entretanto, é que não devemos assumir que, como era quase universalmente assumido na literatura mais antiga, a tese de que todo evento tem uma causa é equivalente à tese do determinismo. Esse é um ponto importante, porque alguns dos antigos argumentos na literatura **contra** o determinismo assumiam que as duas teses são equivalentes (HOBART, 1934).

Na literatura mais antiga, assumia-se que o determinismo é a hipótese com a qual a ciência trabalha, e que rejeitar o determinismo significa negar a ciência, o que não parece plausível. Algumas pessoas pensam que a física quântica mostrou que o determinismo é falso. Isto ainda é controverso (ALBERT, 1992; LOEWER, 1996b; P. LEWIS, 2016), mas, contemporaneamente, é bem aceito que podemos rejeitar o determinismo sem aceitar a visão de que o comportamento de seres humanos se encontra fora do escopo das leis da natureza. Se o naturalismo é a tese de que o comportamento humano pode ser explicado do mesmo modo que as outras coisas do universo, em termos de eventos, processos naturais e leis da natureza, então, podemos rejeitar o determinismo sem rejeitar o naturalismo.

Perceba, por fim, que o determinismo não implica o fisicalismo, nem é implicado pelo fisicalismo. Há mundos possíveis nos quais o determinismo é verdadeiro e o fisicalismo é falso, por exemplo, mundos nos quais as mentes são coisas não-físicas que, mesmo assim, obedecem a leis determinísticas (VAN

INWAGEN, 1998). E há mundos possíveis, talvez o nosso próprio, em que o fisicalismo é verdadeiro e o determinismo é falso.

Já basta de determinismo. E sobre o livre-arbítrio? Como deveríamos entender o desacordo entre compatibilistas e incompatibilistas?

Há muito espaço para debater como, exatamente, deveríamos entender nossas crenças de senso-comum sobre nós mesmos como pessoas dotadas de livre-arbítrio. (Nascemos com livre-arbítrio? Se não, quando o adquirimos? E em virtude de quais capacidades e poderes o temos? Qual a diferença entre agir intencionalmente e agir com livre-arbítrio?). Felizmente, não temos que responder a essas perguntas para dizer o que está sob disputa entre compatibilistas e incompatibilistas. Vamos definir **a tese do livre-arbítrio** como a de que pelo menos uma criatura não-divina (humana) tem livre-arbítrio, deixando em aberto o que isso significa. A tese do livre-arbítrio é uma tese **mínima** sobre o livre-arbítrio; ela seria verdadeira se apenas uma pessoa no universo agisse com livre-arbítrio, ou seja, livremente, possuindo livre-arbítrio em uma ocasião. Não vamos assumir que a tese do livre-arbítrio seja verdadeira, ou mesmo possivelmente verdadeira, mas vamos chamar de **mundo com livre-arbítrio** qualquer mundo possível no qual a tese do livre-arbítrio é verdadeira. Dado que o não-determinismo é a negação do determinismo, e dado que o determinismo é uma tese contingente, podemos dividir o conjunto de mundos possíveis em dois subconjuntos disjuntos: mundos determinísticos e mundos não-determinísticos.

Com este aparato, podemos definir o incompatibilismo e o compatibilismo do seguinte modo: o incompatibilismo é a tese de que **nenhum mundo determinístico é um mundo com livre-arbítrio** (de maneira equivalente, o incompatibilismo é a tese de que, necessariamente, se o determinismo é verdadeiro, então a tese do livre-arbítrio é falsa). E podemos definir o compatibilismo como a negação do incompatibilismo, isto é, como a tese de que alguns mundos determinísticos são mundos com livre-arbítrio (de maneira equivalente, o compatibilismo é a tese de que, possivelmente, determinismo e a tese do livre-arbítrio são ambas verdadeiras).

Este modo de definir o compatibilismo não traz problemas. Há compatibilistas que são agnósticos sobre a verdade ou a falsidade do determinismo, então, um compatibilista não precisa ser um determinista fraco, ou seja, alguém que acredita que o determinismo é verdadeiro e que temos livre-arbítrio. E um compatibilista pode acreditar que não temos livre-arbítrio por razões independentes do determinismo. Mas todos os compatibilistas acreditam que é, ao menos possível, que o determinismo seja verdadeiro e tenhamos livre-arbítrio. Então, todos os compatibilistas estão comprometidos com a tese de que há mundos determinísticos que são mundos com livre-arbítrio.

Porém, esta definição de incompatibilismo tem uma consequência surpreendente. Suponha, como alguns filósofos e filósofas defenderam, que não temos livre-arbítrio porque o livre-arbítrio é conceitualmente ou metafisicamente impossível, pelo menos para seres não-divinos como nós (TAYLOR, 1962, 1992; G. STRAWSON ,1986, 1994). Se estes filósofos estão certos, não há mundos com livre-arbítrio, sendo assim, não há mundos deterministicos com livre-arbítrio. Logo, se o livre-arbítrio é conceitualmente ou metafisicamente impossível, pelo menos para criaturas como nós, segue-se que o incompatibilismo, como o acabamos de definir, é verdadeiro. Mas isto não parece correto. Se é conceitualmente ou metafisicamente impossível que tenhamos livre-arbítrio, então não temos livre-arbítrio, **independentemente**, do determinismo ser verdadeiro ou falso. E, se esse for o caso, então o incompatibilista não pode dizer o tipo de coisas que ele tradicionalmente queria dizer: que a verdade ou falsidade do determinismo é **relevante** para a questão de se temos ou não livre-arbítrio, que, se o determinismo fosse verdadeiro, então não teríamos livre-arbítrio **por causa** do determinismo ser verdadeiro, e assim por diante.

Se desejamos evitar esse resultado contra-intuitivo, há uma solução. Em vez de entender o compatibilismo e o incompatibilismo como proposições que são contraditórias, podemos entendê-las como proposições que são contrárias. Isto é, podemos entendê-las como teses que não podem ser ambas verdadeiras, mas que podem ser ambas falsas. O compatibilismo e o incompatibilismo são ambos falsos se uma terceira tese, o impossibilismo, for verdadeira. O impossibilismo é a tese de que o livre-arbítrio é conceitualmente ou metafisicamente impossível para criaturas não-divinas como nós.

Se aceitarmos essa classificação triádica, podemos definir nossos termos do seguinte modo: Impossibilismo é a tese de que não há mundos com livre-arbítrio, Incompatibilismo é a tese de que há mundos com livre-arbítrio, mas que não há mundos deterministicos que sejam mundos com livre-arbítrio, já o Compatibilismo é a tese de que há mundos com livre-arbítrio e que mundos com livre-arbítrio incluem mundos deterministicos. Para algumas objeções a esta classificação triádica, consulte McKenna (2010), Mickelson (2015a). Para uma defesa, confira Vihvelin (2008, 2013).

O termo “impossibilismo” está sendo cunhado, entretanto, a posição descrita é reconhecida na literatura por uma variedade de nomes: a visão “sem livre-arbítrio de qualquer jeito”, “não-realismo”, “ilusionismo”, “pessimismo”. Teóricos que defendem o impossibilismo incluem Double (1991), G. Strawson (1986, 1994) e Smilansky (2000). Outro tipo de impossibilismo é o fatalismo (TAYLOR, 1962, 1992).

Na literatura mais antiga, havia apenas dois tipos de incompatibilistas, deterministas fortes e libertários. Um determinista forte é um incompatibilista que acredita que o determinismo é verdadeiro (ou, talvez, está próximo o bastante de ser verdadeiro para aquilo que nos interessa, dos modos relevantes para o livre-arbítrio) e, por causa disso, não temos livre-arbítrio (HOLBACH, 1770; WEGNER, 2003). Um libertário é um incompatibilista que acredita que, de fato, temos livre-arbítrio e que isso implica a falsidade do determinismo, de modo relevante (VAN INWAGEN, 1983). Tradicionalmente, libertários acreditam que "o modo relevante" requer que agentes tenham um poder causal especial e misterioso não possuído por mais nada na natureza: um poder quase divino de ser uma causa não-causada de mudanças no mundo (CHISHOLM, 1964). Libertários que aceitam essa posição estão comprometidos, aparentemente, com a tese de que o livre-arbítrio é possível apenas em mundos que são pelo menos parcialmente sem leis e que nosso mundo é um desses mundos (*vide* CLARKE, 2003; O'CONNOR, 2000). Porém, na literatura contemporânea, há incompatibilistas que evitam teses metafísicas tão arriscadas, argumentando que o livre-arbítrio é possível em mundos em que algumas de nossas ações são causadas por eventos indeterminísticos (KANE, 1989, 1996, 1999, 2008, 2011a; EKSTROM, 2000; BALAGUER, 2010) ou que o livre-arbítrio é possível em mundos em que algumas de nossas ações não possuem causas (GINET, 1990). Note que nenhum desses três tipos de incompatibilistas (teóricos da causação do agente, teóricos da causação por eventos indeterminísticos, teóricos não-causais) precisam ser libertários. Eles e elas podem suspender o juízo sobre se o determinismo é verdadeiro ou não e, consequentemente, suspender o juízo sobre se temos ou não livre-arbítrio. Também podem ser deterministas persistentes porque acreditam que o determinismo é verdadeiro. Mas o que todos e todas acreditam – aquilo que os(as) faz incompatibilistas – é que é **possível** termos livre-arbítrio e que ele depende de um **fato contingente** sobre as **leis** que governam o universo: que elas sejam indeterminísticas do jeito certo. Confira o verbete sobre **teorias incompatibilistas de livre-arbítrio**.

Com essas definições e distinções, podemos agora dar o primeiro passo na direção de clarificar o desacordo entre compatibilistas e incompatibilistas. Ambos os lados concordam que é conceitualmente e metafisicamente possível termos livre-arbítrio, e o desacordo entre eles é sobre se algum dos mundos possíveis em que temos livre-arbítrio é um mundo determinístico. O compatibilista diz "sim", já o incompatibilista diz "não". Os argumentos a favor do incompatibilismo precisam, então, ser argumentos a favor da tese de que, necessariamente, se o determinismo

é verdadeiro, não temos o livre-arbítrio que, em outro caso, poderíamos ter.

2. Duas Razões para Pensar que o Livre-arbítrio é Incompatível com o Determinismo

Uma reação inicial ao determinismo consiste em pensar que nossas escolhas **não fazem diferença alguma** sobre o que acontece, já que há causas anteriores determinaram previamente ou "fixaram" todo o futuro (NAHMIAS, 2011). É fácil pensar que o determinismo implica que temos um destino que não podemos evitar, não importa o que escolhemos ou decidimos e não importa o quanto tentamos.

> O homem, frequentemente sem saber, frequentemente a despeito de si, quando confronta o caminho que a natureza lhe definiu, lembra um nadador que é obrigado a seguir a correnteza que o carrega; ele acredita ser um agente livre porque às vezes consente, às vezes não consente, em nadar com a correnteza, a qual, não obstante, sempre o leva para frente (HOLBACH, 2002, p. 181). [1]

É amplamente aceito, tanto por incompatibilistas quanto por compatibilistas, que este pensamento é equívoco. As descobertas empíricas sobre o cérebro e o comportamento humanos podem nos informar que não temos tanto controle consciente quanto pensamos ter (WEGNER, 2003; LIBET, 1999). Para críticas dos argumentos defendendo que pesquisas científicas recentes mostraram que “a vontade consciente é uma ilusão”, veja Mele (2009), alguns dos ensaios em Sinnott-Armstrong e Nadei (2011) e Roskies e Nahmias (2016). E há também preocupações, levantadas por certas versões do fisicalismo, de que nossos estados mentais não possuem os poderes causais que pensamos possuírem (KIM, 1998). Mas essas ameaças ao livre-arbítrio não têm a ver com o determinismo. Ele pode implicar que nossas escolhas e esforços têm causas anteriores suficientes, mas não implica que não temos o poder de escolha ou que nossas escolhas e esforços sejam causalmente impotentes. O determinismo é consistente com o fato de que nossa deliberação, nossas escolhas e nossos esforços fazem parte do processo causal pelo qual

---

[1] Confira, também, Wegner (2003).

nossos corpos se movem, e, também, causam outros efeitos no mundo. E uma causa é o tipo de coisa que "faz uma diferença" (SARTORIO, 2005). Se levanto minha mão porque o escolho fazer, então é verdade, *ceteris paribus*, que, se eu não tivesse assim escolhido, o evento de levantar a minha mão não teria ocorrido.

Colocando de lado esta preocupação, podemos classificar os argumentos a favor do incompatibilismo a partir da seguinte caracterização:

1. Os argumentos a favor da tese de que o determinismo tornaria impossível que **causássemos e controlássemos nossas ações da maneira correta**.

2. Argumentos a favor da tese de que o determinismo nos tiraria **o poder ou a capacidade para agir ou escolher de outro modo**.

Os argumentos do primeiro tipo costumam focar nas noções de Eu, causação e responsabilidade; a preocupação aqui é que o determinismo elimina o tipo de causação que evocamos quando atribuímos ações às pessoas ("Foi a Maria que quebrou o vaso") e fazemos julgamentos sobre a responsabilidade moral ("Não foi culpa dela. Cíntia a empurrou."). Alguém que defende o incompatibilismo desse jeito pode admitir que a verdade do determinismo seja consistente com a eficiência causal das nossas deliberações, escolhas e tentativas de ação. Porém, esta pessoa insistiria que o determinismo implica que o **único** sentido em que somos responsáveis pelo que fazemos é o qual um cachorro ou uma criança pequena é responsável. Nesta visão, a responsabilidade moral requer algo a mais que isso. A responsabilidade moral requer autonomia ou autodeterminação: que nossas ações sejam causadas e controladas **apenas** por nós mesmos. Para usar o *slogan* popular na literatura: agimos livremente e somos moralmente responsáveis apenas quando **somos** a origem última de nossas ações.

> Cada um de nós, quando agimos, é um primeiro motor imóvel. Fazendo o que fazemos, causamos o acontecimento de certos eventos, e nada – nem ninguém - nos causa causar o acontecimento desses eventos. (CHISHOLM, 1964, p. 32).

> O livre-arbítrio... é o poder de agentes de serem os criadores últimos ou originadores e sustentadores

> dos próprios fins e propósitos... quando traçamos as cadeias causal ou explanatória de ações até suas origens nos propósitos de agentes livres, essas cadeias causais devem chegar ao fim ou terminar nas volições (escolhas, decisões, ou esforços) de agentes, que causam ou tornam realidade seus propósitos. (KANE, 1996, p. 4).

Os argumentos do segundo tipo focam na noção de escolha. Ter uma escolha, aparentemente, é ter opções ou alternativas genuínas - diferentes modos em que podemos agir. A preocupação é que o determinismo implica que aquilo que fazemos é, sempre, a **única** coisa que podemos fazer, e, por causa disso, nunca **realmente** temos uma escolha **sobre algo**, em oposição a estar sob efeito da (talvez inescapável) ilusão de que temos uma escolha. Alguém que defende o incompatibilismo desse modo pode conceder que a verdade do determinismo é consistente com nossas escolhas, pelo menos no sentido em que um cachorro ou uma criança pequena faz escolhas, e consistente também com nossas escolhas serem causalmente eficientes. Mas, esta pessoa insiste, isso não é o bastante para termos livre-arbítrio; temos livre-arbítrio apenas se temos uma escolha genuína sobre quais ações executamos, e temos uma escolha genuína somente se há mais de uma ação que **somos capazes** de executar.

> Uma pessoa tem livre-arbítrio se ela frequentemente se encontra em posições como as seguintes: ela deve agora falar ou ficar em silêncio, e *pode* falar agora e *pode* permanecer em silêncio; deve tentar salvar a criança se afogando ou então buscar ajuda, e é *capaz* de tentar resgatar a criança e *capaz* de buscar ajuda; deve agora resignar da presidência ou então mentir para os membros; e está no alcance de seu poder resignar e está no alcance de seu poder mentir. (VAN INWAGEN, 1983, p. 8).

Van Inwagen está dando o que ele considera que sejam exemplos paradigmáticos de pessoas que têm livre-arbítrio; ele não está dizendo que o livre-arbítrio **seja** apenas a capacidade de executar mais do que uma ação explícita. Nossas escolhas incluem escolhas entre ações puramente mentais (prestar atenção

numa aula ou passar o resto do tempo decidindo o que cozinhar na janta) assim como escolhas sobre as ações que executamos movendo nossos corpos. Se o incompatibilista afirmar que o determinismo nos tira o livre-arbítrio porque nos impede de escolher, isto deve ser entendido como a tese de que o determinismo tem a consequência de que nunca somos capazes de fazer **algo** que não realmente fazemos, em que "fazer" inclui decidir, escolher, e outras instâncias de agência mental, assim como coisas que fazemos movendo nossos corpos.

Podemos nos perguntar se os argumentos baseados na autodeterminação e os argumentos baseados na escolha são modos independentes de argumentar a favor do incompatibilismo pela seguinte razão: eu causo e controlo minhas ações do modo autodeterminante, tal como requerido para se ter responsabilidade moral, somente se minhas ações são o produto do meu livre-arbítrio, e minhas ações são ações do meu livre-arbítrio somente se tenho a capacidade de agir (escolher agir, decidir agir, intencionar agir, tentar agir) de outro modo. Se o determinismo tem a consequência de que **nunca** tenho a capacidade de agir de outro modo, também tem a consequência de que nunca causo as minhas ações do modo autodeterminante requerido para se ter responsabilidade moral (KANE, 1996).

Havia uma época em que a ligação entre responsabilidade moral, autodeterminação e a capacidade para agir de outro modo era um terreno comum entre compatibilistas e incompatibilistas. Isto é, todos e todas aceitavam que uma pessoa é moralmente responsável somente se ela tem o tipo certo de **controle** sobre o que ela faz, e todos pressupunham que uma pessoa tem o tipo correto de controle sobre algo que ela faz somente se ela é **capaz de agir** *(***ou***,* **pelo menos, decidir, escolher, intencionar ou tentar***)* **de outro modo**. Dado esse pressuposto, qualquer pessoa com a pretensão de defender a tese de que a responsabilidade moral é compatível com o determinismo tinha primeiro que mostrar que a capacidade de agir de outro modo é compatível com o determinismo. Durante o auge da filosofia da linguagem comum na metade do século passado, havia um pressuposto adicional: que a capacidade para agir de outro modo é compatível com o determinismo somente se sentenças como "*S* poderia ter feito ***X*"** (em que **X** é algo que **S** não fez) fossem equivalentes em significado a condicionais contrafactuais como "Se **S** tivesse escolhido (ou tentando, desejado, preferido, etc.) fazer **X**, **S** teria feito **X"**. Esse debate foi prejudicado por se passar numa época em que contrafactuais não eram bem entendidos, antes do advento das semânticas de mundos possíveis de Lewis-Stalnaker (D. LEWIS, 1973). Havia contraexemplos às análises propostas e havia um consenso crescente de que os prospectos para uma "Análise Condicional" de

sucesso eram baixos (AUSTIN, 1961; LEHRER, 1976). Para uma defesa de que esse pessimismo era prematuro, confira Vihvelin (2004, 2013). Para uma defesa de que um compatibilista não precisa defender uma Análise Condicional, consulte Lehrer (1976).

Foi nesse contexto que Harry Frankfurt propôs seu famoso contraexemplo ao "Princípio de Possibilidades Alternativas" (uma pessoa é moralmente responsável pelo que ela fez somente se ela pudesse agir de outro modo). Frankfurt queria defender a tese de que a responsabilidade moral é compatível com o determinismo sem ter que defender a tese de que a capacidade de agir de outro modo é compatível com o determinismo. Sua estratégia foi usar um engenhoso experimento mental que deveria mostrar que não importa como você entende a capacidade de agir de outro modo, se você é um compatibilista ou incompatibilista, você deve aceitar que possuir esta capacidade não é uma condição necessária para ser moralmente responsável (FRANKFURT, 1969).

Havia duas etapas para o experimento mental. Na primeira etapa, Frankfurt nos convida a imaginar uma pessoa, Jones, que tem livre-arbítrio, e que escolhe agir livremente e satisfaz todas as condições que você considera necessárias e suficientes para a responsabilidade moral. Você pode imaginar Jones em algum dos cenários descritos por van Inwagen, diante de uma escolha de falar ou ficar em silêncio, de tentar resgatar a criança ou buscar ajuda, de resignar a presidência ou mentir, e imaginar Jones deliberando e decidindo, por suas próprias razões, em favor de uma das alternativas contempladas, e então agindo com sucesso a partir de sua decisão. Na segunda etapa, você é convidado a adicionar à história a existência de um ser poderoso, Black, que tem grande interesse no que Jones faz, incluindo como delibera e decide. Você pode preencher os detalhes como preferir, mas você deve imaginar que Black tenha o poder de interferir em Jones de tal modo a assegurar que Jones faça **exatamente** o que Black deseja que ele faça. E você deve imaginar também que Black está muito atento e está preparado a intervir instantaneamente, caso seja necessário, para impedir Jones de fazer algo que Black não deseja que ele faça. Porém, no final das contas, nada disso foi necessário. Por pura e fortuita coincidência, Jones fez exatamente o que Black desejava que ele fizesse (ele até mesmo deliberou e decidiu do jeito que Black queria que ele deliberasse e decidisse). Então Black permaneceu à espreita apenas assistindo a tudo.

Já que Black nada fez em relação a Jones, parece que Jones é tão moralmente responsável na segunda etapa da história quanto o é na primeira etapa. Mas também parece que a existência do poderoso Black seja suficiente para fazer

com que seja o caso que Jones é incapaz de agir de outro modo (ou, como Frankfurt coloca, ele não tem “possibilidades alternativas”).

A história de Frankfurt tem um forte apelo intuitivo, e muitos se convenceram que ela mostra que estávamos errados quando pensávamos que uma pessoa tinha o tipo correto de controle sobre o que ela fazia, o tipo necessário para a responsabilidade moral, somente se ela era capaz de agir de outro modo. Esses filósofos e essas filósofas, que mais tarde ficaram conhecidos(as) como compatibilistas “de origem” ou “semi-compatibilistas”, acharam que a lição a ser aprendida era que não é relevante, **enquanto a preocupação for com a responsabilidade moral**, se o determinismo tem como consequência que não temos o livre-arbítrio que pensamos ter, que nunca temos qualquer escolha genuína, que nunca somos capazes de agir de outro modo. Eles(as) dizem que o que importa, enquanto a preocupação for com responsabilidade moral, é apenas o que acontece na “sequência atual” (FISCHER, 1994; FISCHER; RAVIZZA, 1998; MCKENNA, 2001; SARTORIO, 2016).

Mas um pouco de reflexão deve nos fazer pensar se fatos sobre a “sequência atual” podem ser separados de fatos sobre o livre-arbítrio do modo almejado por Frankfurt. Como pode ser verdade que Black obtém sucesso não apenas em limitar o **escopo** de cursos alternativos de ação de Jones, mas em fazer com que seja o caso que Jones **não** tenha alternativas? Como Black pode, ficando à espreita, tirar de Jones a capacidade de deliberar, decidir ou agir de outro modo?

Podemos facilmente conceder que Black tem o poder de manipular diretamente o corpo de Jones de modo que Jones perca a capacidade de mover seu corpo de qualquer outro modo que não o modo desejado por Black. Podemos até mesmo aceitar que Black tem o poder de manipular diretamente a mente/cérebro de Jones, de tal modo que Jones perca a capacidade de “mover” sua mente de qualquer modo que não o modo desejado por Black. Mas Black nunca exercita esse poder. Há uma diferença entre a **existência** de um poder e o **exercício** de um poder. É uma falácia modal concluir a partir de "Black tem o poder de fazer com que Jones não tenha a capacidade de agir de outro modo" que "Jones não tem a capacidade de agir de outro modo". A verdade sobre Jones não é que Black tira sua capacidade para agir de outro modo; é a verdade mais complicada de que Black o coloca em constante **risco** de perder a capacidade de agir de outro modo. O livre-arbítrio de Jones e sua responsabilidade moral dependem da sorte de suas deliberações e intenções coincidirem com as de Black (VIHVELIN, 2013). Para algumas outras críticas ao argumento de Frankfurt, veja Lamb (1993), Alvarez (2009) e Steward (2009).

O argumento de Frankfurt foi extremamente influente, ainda que talvez não do jeito que o autor pretendia. Seu experimento mental foi um fracasso, enquanto a maioria dos compatibilistas foi convencida, a maioria dos incompatibilsitas não foi. Os compatibilistas que não se convenceram incluem Smith (1997, 2004), Campbell (2005), Fara (2008) e Vihvelin (2013). Esses(as) incompatibilistas insistem, ainda que não pela razão exposta anteriormente, que Black não consegue eliminar de Jones toda a sua liberdade, ainda há **algo** que Jones possui (WIDERKER, 1995; GINET, 1996; KANE, 1996). Esta resposta, apelidada famosamente de resposta dos "lampejos de liberdade", foi criticada por defensores do argumento de Frankfurt com base na ideia de que as "possibilidades alternativas" mantidas por Jones não são suficientemente "robustas" (FISCHER, 1994, 2003, 2011; FISCHER; RAVIZZA, 1998). Os críticos do argumento rejeitam essa objeção, defendendo que Jones mantém uma capacidade de agir de outro modo moralmente relevante, ressuscitando assim o mesmo debate que Frankfurt esperava desqualificar. Outros defensores do argumento de Frankfurt contaram histórias mais complicadas e argumentaram que, mesmo se a história de Frankfurt não obtiver sucesso, essas novas narrativas mostrariam que uma pessoa pode ser moralmente responsável por sua ação ainda que **não** tenha **qualquer** capacidade para agir de outro modo (MELE; ROBB, 1998; PEREBOOM, 2003). Quase cinquenta anos depois da publicação do artigo de Frankfurt, o debate ainda continua. Para uma amostra de parte desta vasta literatura, confira Widerker e McKenna (2003).

O objetivo de Frankfurt era tornar mais fácil defender a tese de que a **responsabilidade moral** é compatível com o determinismo, e pode-se dizer que ele obteve sucesso ou, pelo menos, sucesso contra os incompatibilistas "de Origem", que ficaram convencidos pelo seu argumento (FISCHER, 2003). Mas houve um custo. Nosso interesse no livre-arbítrio não é limitado pelo nosso interesse na responsabilidade moral. Historicamente, acusou-se o compatibilismo de ser tão somente um "infeliz subterfúgio" ou um "lamaçal da evasividade". O determinismo parece, *prima facie*, incompatível com escolhas genuínas e a capacidade de agir de outro modo, e ainda que, por um tempo, o consenso filosófico era que o incompatibilista incorria em alguma confusão ou erro parecido com o erro fatalista, este não é mais o caso, graças ao Argumento da Consequência (que será discutido logo mais em seguida). A influência do argumento de Frankfurt, e de seus outros ensaios, em especial, Frankfurt (1971), foi tão grande que a maior parte dos compatibilistas virou as costas para o problema do livre-arbítrio/determinismo e focou exclusivamente nos problemas sobre a responsabilidade moral (e tópicos

relacionados, como culpa, merecimento e punição). Atualmente, usa-se o termo "compatibilismo" mais comumente para se referir a uma tese sobre a responsabilidade moral – a liberdade que é suficiente para a responsabilidade moral é compatível com o determinismo (PEREBOOM, 1995), a existência de ações livres no sentido requerido para se ter responsabilidade moral é compatível com o determinismo (MARKOSIAN, 1999), a capacidade única de pessoas de exercer controle da maneira necessária para a responsabilidade moral é compatível com o determinismo (MCKENNA, 2004b). A literatura sobre o problema tradicional do livre-arbítrio e do determinismo é dominada por incompatibilistas. Há um consenso crescente de que o incompatibilista está correto: se nosso universo for determinístico, nunca teremos a capacidade de escolher e agir de modo diferente daquele que de fato agimos.

Antes de nos perguntarmos se este pessimismo em relação à compatibilidade entre o livre-arbítrio e o determinismo é bem fundado, devemos parar um momento para perguntar se realmente há um desacordo substancial entre compatibilistas e incompatibilistas. Quando um incompatibilista diz que o determinismo nos rouba o livre-arbítrio que pensamos ter, incluindo escolhas genuínas e a capacidade de agir de outro modo, e quando o compatibilista nega isso, eles estão afirmando e negando a mesma proposição? Ou o incompatibilista está afirmando uma coisa e o compatibilista negando outra?

Algumas passagens na literatura sugerem que não há um debate substantivo. Por exemplo, alguns filósofos e algumas filósofas contrastam a "capacidade incompatibilista forte" com a "capacidade compatibilista fraca" (ou o "livre-arbítrio libertário" com o "livre-arbítrio compatibilista") e escrevem de um modo que sugere que elas pensam que a única questão substantiva sobre livre-arbítrio é se ele é, ou implica, a capacidade incompatibilista "forte" de agir de outro modo (já que a capacidade compatibilista "fraca" de agir de outro modo já é, por definição, compatível com o determinismo). E uma proposta semântica popular pode parecer apoiar a tese de que não há uma disputa real. De acordo com a proposta Lewis/Kratzer, "poder" (e outros termos modais, incluindo "é capaz", "tem o poder para", "é livre para") significam "co-possível com os fatos relevantes ***F'***", e os fatos relevantes são determinados pelo contexto junto das intenções do falante (D. LEWIS, 1976; KRATZER, 1977). Para uma proposta contextualista diferente, consulte Hawthorne (2001), e para críticas, Feldman (2004).

Então, por exemplo, pode ser verdade, em um contexto, que eu posso falar finlandês (eu sou uma falante nativa de estônio, então não levaria muito tempo para que eu aprendesse finlandês) e falso em outro contexto (não me leve para a Finlândia

como sua intérprete). Não há contradição em dizer que sou tanto capaz quanto incapaz de falar finlandês, desde que entendamos que o que estamos dizendo é que eu falar finlandês é co-possível com um conjunto de fatos, como minha capacidade de facilmente aprender a língua, e não é co-possível com outro conjunto de fatos mais inclusivo, como minha atual incapacidade de falar finlandês. Dada essa interpretação de "poder", pode parecer que não há um desacordo genuíno entre compatibilistas e incompatibilistas. Quando o(a) compatibilista afirma que agentes determinísticos frequentemente têm a capacidade de agir de outro modo, ele(a) tem em mente contextos (os contextos de falantes normais, do dia a dia) em que os fatos relevantes **F** são restritos a alguns **fatos relativamente locais** sobre o agente e seu arredor. Quando o(a) incompatibilista nega que um agente determinístico é capaz de agir de outro modo, ele(a) tem em mente contextos (os contextos dos debates filosóficos sobre o problema do livre-arbítrio/determinismo) em que os fatos relevantes incluem **todos os fatos sobre as leis da natureza e o passado**. Logo, a proposição negada pelo incompatibilismo não é a proposição afirmada pelo compatibilismo.

O principal incompatibilista contemporâneo, van Inwagen, rejeita qualquer sugestão de que o compatibilista e o incompatibilista tem um desacordo meramente verbal, seja porque eles estejam usando diferentes sentidos de "poder" ("capacidade", "livre-arbítrio", etc.) ou porque eles estão focando em diferentes contextos de asserção. O debate, ele diz, é sobre se o determinismo tem como consequência que ninguém é capaz de agir de outro modo (de maneira equivalente, não há uma pessoa que tenha sob seu poder a possibilidade de agir de outro modo) dado o que **falantes normais** querem dizer nos contextos em que eles usam esses termos. O contexto a que ele se refere são os contextos de deliberação e escolha em que consideramos nossas opções enquanto acreditamos que somos capazes de seguir cada uma delas; Quando um compatibilista afirma, e um incompatibilista nega, que uma pessoa em um mundo determinístico às vezes é capaz de agir de outro modo, eles querem dizer a mesma coisa com "ser capaz de", eles querem dizer o que falantes de inglês querem dizer (nesses contextos). A proposição asserida por compatibilistas é a proposição negada por incompatibilistas. Citando David Lewis, como seu exemplo de um oponente compatibilista, van Inwagen diz que ele e Lewis não podem ambos estar certos. Um deles está errado, mas nenhum deles está confuso ou cometendo um erro simples (VAN INWANGEN, 2008).

No que se segue, assumiremos que o debate sobre o livre-arbítrio, incluindo, mas não necessariamente limitado a escolhas genuínas e a capacidade para agir de outro modo, e o determinismo é um debate substantivo, e não um debate que

pode ser dissolvido apelando a diferentes sentidos ou contextos de proferimento. Concentremo-nos então nos argumentos.

3. Os argumentos baseados em Intuição

Alguns argumentos a favor do incompatibilismo não se encaixam nos dois tipos descritos anteriormente – os argumentos a favor da tese de que o determinismo é incompatível com a noção de origem última e os argumentos a favor da tese de que o determinismo é incompatível com a capacidade de agir de outro modo – pois são argumentos que apelam primariamente a nossas intuições. Há várias maneiras de se argumentar assim a favor do incompatibilismo, mas a estrutura básica deste argumento é normalmente a seguinte:

Se o determinismo for verdadeiro, somos como: bolas de bilhar, brinquedos de corda, joguetes de forças externas, marionetes, robôs, vítimas de um nefasto neurocirurgião que nos controla através da manipulação direta dos estados cerebrais que são a causa imediata de nossas ações. Bolas de bilhar (brinquedos de corda, etc.) não têm livre-arbítrio. Portanto, se o determinismo for verdadeiro, não temos livre-arbítrio.

Mas os argumentos baseados em intuição geralmente não são muito bons. Mesmo que o determinismo implique que temos **algo** em comum com coisas que não têm livre-arbítrio, disso não se segue que não há diferenças relevantes. Bolas de bilhar, brinquedos, marionetes e robôs simples não possuem uma mente, e ter uma mente é condição necessária para ter livre-arbítrio. E o determinismo não tem como consequência que todas as nossas ações são causadas por desejos irresistíveis que são, como as manipulações diretas do neurocirurgião, impostas a nós por forças externas fora do nosso controle. Para críticas às "bombas de intuição", veja Dennet (1984). Para discussões dos casos envolvendo manipulações mais discretas, confira a Seção 3.2.

Estando atentos a isso, vejamos os dois argumentos baseados em intuição mais influentes.

3.1. O Argumento da Ausência de Bifurcações

O Argumento da Ausência de Bifurcações (VAN INWAGEN, 1983; FISCHER, 1994; EKSTROM, 2000) começa apelando à ideia de que, quando fazemos uma

escolha, fazemos algo (ou pensamos que fazemos algo tal como um viajante o faz, quando ele se depara com uma escolha entre duas estradas diferentes. As únicas estradas que um viajante é capaz de escolher são estradas que são um contínuo da estrada em que ele já está. Analogamente, as únicas escolhas que somos capaz de realizar são escolhas que são contínuos **do passado atual e consistentes com as leis da natureza**. Se o determinismo for falso, então fazer escolhas realmente é desse jeito: uma "estrada" (o passado) atrás de nós, duas ou mais "estradas" diferentes (ações futuras consistentes com o passado) à nossa frente. Mas se o determinismo for verdadeiro, então nossa jornada pela vida é como viajar (em apenas uma direção) em uma estrada sem bifurcações; há outras estradas, levando a outros destinos; se pudéssemos ir a uma dessas outras estradas, poderíamos chegar a outro destino. Mas não conseguimos chegar a qualquer dessas outras estradas a partir da estrada em que estamos **realmente**. Logo, se o determinismo for verdadeiro, nosso futuro atual é *o* **único futuro possível** para nós; nunca somos capazes de escolher ou fazer nada diferente daquilo que realmente fazemos. Confira, também, Flint (1987) e Warfield (2003), para uma discussão de um argumento relacionado que apela à metáfora de que nossa liberdade "se adiciona" à lista de verdades sobre o mundo.

Esta ideia é intuitivamente forte, já que é natural pensar que nosso futuro é "aberto", tal como a analogia da estrada bifurcada sugere, e, também, é natural associar esse tipo de estrutura bifurcada à liberdade de escolha. Mas vários pressupostos cruciais foram aceitos por debaixo dos panos nesse cenário: pressupostos sobre o tempo e a causação e pressupostos sobre o que é possível. Os pressupostos sobre o tempo e a causação necessários para a analogia funcionar são os seguintes: que nós nos "movemos" através do tempo de modo parecido com o modo como nos movemos numa estrada, que nosso "movimento" é, necessariamente, apenas em uma direção, do passado para o futuro, que o passado é, **necessariamente**, "fixo" ou fora do nosso controle de um jeito que o futuro não é. Todos esses pressupostos são controversos, pois em algumas teorias de tempo e causação, a quadri-dimensionalista do tempo, que permite viagem no tempo e causação retroativa, eles são todos falsos (D. LEWIS, 1976; HORWICH, 1987; SIDER, 2001; HOEFER, 2002).

O pressuposto sobre o que é possível é este: mundos possíveis são entidades espaço-temporais concretas (do modo como estradas o são) e que mundos possíveis podem se sobrepor (literalmente, dividir uma parte comum) do modo como estradas podem se sobrepor. Mas a maior parte dos teóricos e teóricas de mundos possíveis rejeitam o primeiro pressuposto e praticamente todo o mundo

rejeita o segundo pressuposto (ADAMS, 1974; D. LEWIS, 1986).

O determinismo, sem esses pressupostos adicionais, **não** tem como consequência que nossa "jornada" pela vida seja como seguir uma estrada; o contraste entre o determinismo e o não-determinismo **não** é o contraste entre viajar em uma estrada com bifurcações e viajar em uma estrada sem bifurcações.

Como um argumento a favor do incompatibilismo, o apelo da metáfora das estradas bifurcadas ("O jardim dos caminhos bifurcados") falha. Se tirarmos as metáforas, a premissa principal do argumento se torna a tese de que temos escolhas genuínas entre diferentes cursos de ação somente se o poder para **escolher e agir de outro modo for co-possível com o passado atual e as leis atuais**. Mas esta é precisamente a tese que está em disputa entre compatibilistas e incompatibilistas.

Se as intuições a que o Argumento da Ausência de Bifurcações apela ainda assim continuam a nos engajar, é porque pensamos que o nosso escopo de escolhas possíveis é limitado por dois fatores: as leis e o passado. Não podemos mudar ou violar as leis, não podemos afetar o passado (mesmo se a causação retroativa for logicamente possível, não se encontra em **nosso** alcance). Essas crenças – sobre as leis e o passado – são a base para o argumento contemporâneo mais influente a favor do incompatibilismo: o Argumento da Consequência. Mais sobre isto em breve.

3.2. Argumentos da Manipulação

Voltemos nossa atenção agora a uma família de argumentos que apelam à nossa resposta intuitiva a casos envolvendo duas pessoas, que chamaremos de "Vítima" e "Produtor". O Produtor projeta ou manipula a Vítima (em algumas das histórias, do modo como um criador de um robô projeta seu robô ou um deus cria um ser humano; em outras histórias, empregando técnicas de engenharia comportamental ou manipulação neural). O objetivo do Produtor é assegurar ou que a Vítima execute uma determinada ação (MELE, 1995, 2006; ROSEN, 2002; PEREBOOM, 1995, 2001, 2008, 2014) ou que ela terá o tipo de estrutura motivacional e psicológica que garantirá, ou tornará provável, que execute certos tipos de ação e viva um determinado estilo de vida. Consulte Kane (1996) para discussões sobre *Admirável Mundo Novo*, de Huxley, e *Walden II*, de Skinner.

É esperado que concordemos que a Vítima não seja moralmente responsável porque ela não age livremente e que não age livremente por causa do papel desempenhado pelo Produtor na causação de sua ação. A Vítima executou as ações

que executou porque assim foi projetada ou manipulada, e foi o Produtor quem a fez ser desse jeito ou agir desse modo.

O argumento então continua como se segue:

> 1. A Vítima não age livremente e, por esse motivo, não é moralmente responsável pelo que faz.
> 2. Se o determinismo for verdadeiro, não há diferença relevante entre a Vítima e **qualquer** caso normal de uma ação aparentemente livre e moralmente responsável.
> 3. Portanto, se o determinismo for verdadeiro, ninguém nunca age livremente ou é moralmente responsável pelo que faz.

Espera-se que aceitemos a premissa 1 com base na nossa resposta intuitiva à história sobre a Vítima. O argumento pela premissa 2 é que, se o determinismo for verdadeiro, então somos como a Vítima com respeito ao fato de que somos apenas causas **próximas** das nossas ações. Fazemos o que fazemos por causa do jeito que somos (nosso psicológico ou "projeto" junto do conjunto total de nossos pensamentos, desejos e outros estados psicológicos na hora da ação) e as causas dessas características psicológicas **têm origem última fora de nós**, **em forças e fatores fora do nosso controle**. A única diferença entre nós (nesse cenário imaginado em que o determinismo é verdadeiro) e a Vítima é que nossos aspectos psicológicos não são o resultado do trabalho de um único Produtor que tem um plano específico para nós. Mas o fato sobre as causas remotas de nossas ações , que são causadas por uma variedade de causas naturais ao invés das ações intencionais de um único agente, não é relevante para as questões sobre nossa liberdade e responsabilidade, ou, pelo menos, é o que dizem os defensores dos Argumentos de Manipulação.

Os argumentos da Manipulação podem ser vistos por meio da resposta do incompatibilismo de origem ao experimento mental de Frankfurt. Frankfurt tentou mostrar que o determinismo não é tão ruim quanto podemos pensar. Em sua história, Black era um substituto para o determinismo e Frankfurt tentava nos convencer que os fatos sobre Black são consistentes com os fatos, como os conhecemos, sobre como **realmente** deliberamos, decidimos e agimos, e tais fatos são os únicos fatos que importam quando estamos interessados na responsabilidade moral. Então, mesmo que Jones não tenha a habilidade para agir de outro modo, ele ainda é moralmente responsável. O Argumento da Manipulação diz algo do tipo:

> O experimento mental de Frankfurt é falho. O determinismo não é um poderoso agente inerte no fundo, como Black. O determinismo é parte da "sequência atual". Contemos uma história para tornar isso claro [...].

E então o Produtor é apresentado, e nos é dito que ele tem um plano em relação à ação ou ações de outra pessoa, a Vítima, e o poder para concretizar seu plano e, além disso, **diferentemente de Black**, ele o concretiza.

Seria um erro, entretanto, pensar que a manipulação de uma pessoa por outra automaticamente impossibilita a liberdade. Na vida real, sabemos que podemos ser manipulados por outros para fazer coisas que não teríamos feito, mas que seus argumentos ou outros modos de persuasão nos fizeram mudar de ideia. Na ausência de outras razões, não pensamos que esse tipo de manipulação nos tira o livre-arbítrio ou a capacidade de agir livremente do modo requerido para se ter responsabilidade moral. Pensamos que poderíamos ter resistido ao argumento (ou ao papo de vendedor ou às pressões sutis exercidas por nossos amigos e colegas manipuladores) e poderíamos culpar a nós mesmos mais tarde por não termos feito isso.

A questão, então, é saber se há um caso que possa servir aos propósitos de um argumento da manipulação: um caso no qual a Vítima carece da liberdade que é condição necessária para a responsabilidade moral, enquanto isso não é diferente, em qualquer modo relevante, de um agente normal em circunstâncias normais em um mundo determinístico (isto é, de alguém que pensaríamos agir livremente, sendo moralmente responsável pelo que faz).

Há casos e casos, e muitos na literatura não são descritos de maneira satisfatória. Os primeiros três Plums de Pereboom 2001 são um exemplo. Dependendo de como preenchemos os detalhes, nosso veredito sobre Plum pode ser que ele não é livre e não é moralmente responsável porque ele é diferente, em modos relevantes que não requerem a falsidade do determinismo, de um agente normal em circunstâncias normais (FISCHER, 2004; MICKELSON, 2010). Alternativamente, a história pode ser preenchida de um modo que dá suporte para o julgamento de que Plum **não** *é* diferente, em qualquer modo relevante, de um agente normal, determinístico ou indeterminístico, em circunstâncias normais. Mas isso permite ao compatibilista dar uma resposta radical (MCKENNA, 2004a, 2008) que, dado que agentes determinísticos normais são moralmente responsáveis, Plum também o é.

As opiniões variam sobre se o custo intuitivo da resposta radical é muito alto.

Considere agora os casos do tipo *Admirável Mundo Novo*, nos quais as crianças são submetidas à engenharia comportamental intensa desde o nascimento, com a intenção de fazê-las aceitar os papéis que lhe são designados em uma sociedade rigidamente hierárquica. Tudo depende dos detalhes, mas certamente não é implausível pensar que as vítimas de alguns casos do tipo *Admirável Mundo Novo* não têm uma liberdade moralmente significante porque suas capacidades cognitivas, avaliativas e volicionais foram limitadas ou prejudicadas de certos modos:

> Elas são incapazes de efetivamente imaginar ou ver o significado de certas alternativas, de refletir sobre si e as origens de suas motivações. (WATSON, 1987).

Wolf (1990) defende, com base em casos similares, que a habilidade que fundamenta nossa liberdade e responsabilidade é a capacidade não-prejudicada de escolher e agir de acordo com "o Verdadeiro e o Bom" (*vide* NELKIN, 2011). Mas o determinismo não tem como consequência que as capacidades cognitivas, avaliativas e volicionais de todo mundo são prejudicadas deste modo.

Há casos em que a Vítima está sob o controle direto do Produtor, de modo que faz com que seja verdadeiro que a Vítima não é moralmente responsável pelo que faz porque ela não tem mais o tipo de controle causal necessário para agir livremente. O Produtor pode, por exemplo, manipular diretamente os membros da Vítima, de modo que ela mova seu corpo como um fantoche, de uma maneira que a Vítima não gostaria, ou pode manipular diretamente os estados cerebrais, que são os realizadores neurais dos desejos de primeira ordem da Vítima, a fim de que ela, assim como o viciado relutante de Frankfurt, aja intencionalmente mas "contra sua vontade" (FRANKFURT, 1971). Porém, o determinismo não tem como consequência que nunca agimos intencionalmente e que sempre agimos contra nossos desejos de segunda ordem ou, ainda, que sempre estamos submetidos a compulsões psicológicas, vício ou desejos "irresistíveis", fobias e outras aversões patológicas.

Os(as) defensores(as) dos Argumentos da Manipulação afirmam, entretanto, que o argumento funciona mesmo se deixarmos de lado esses tipos de casos. Essas pessoas dizem que a força intuitiva do argumento depende **apenas** do fato de que as ações deterministicamente causadas têm como causa última, assim como as ações da Vítima, fatores e forças fora do controle do agente. Os(as) defensores(as) dos Argumentos da Manipulação dizem que o argumento é bem-sucedido mesmo

em casos nos quais o Produtor projeta a Vítima com capacidades e psicológico normais, talvez do tipo que gostaríamos que nossas crianças tivessem, talvez do tipo psicológico racionalmente egoísta que preferiríamos que nossas crianças não tivessem. Também dizem que o argumento é bem-sucedido, mesmo quando o Produtor é descrito como um projetista tão sofisticado de modo que a Vítima tenha uma história passada que satisfaça todos os requerimentos das teorias compatibilistas da ação livre que incluem uma condição histórica. Para uma boa explicação da diferença entre as teorias compatibilistas históricas e a-históricas no contexto dos Argumentos de Manipulação, consulte McKenna (2004a).

O melhor exemplo deste tipo é o caso de Ernie, apresentado por Mele (2006). Ernie foi criado como um zigoto **Z** em Maria, pela deusa Diana, que desejava que certo evento ocorresse em 30 anos, e a deusa foi capaz de usar seu conhecimento das leis determinísticas e do estado do universo inteiro para deduzir que colocar um zigoto, com a constituição precisa de **Z**, em Maria, produziria um adulto normal (ou melhor que normal: com autocontrole ideal, racional, etc.) que iria, 30 anos mais tarde, julgar, com base em deliberações racionais, que o melhor a se fazer é **A** e, com base neste julgamento racional, que ele faria **A**, de modo a fazer com que **E** ocorresse.

Para muitas pessoas, parece ser intuitivamente claro que Ernie não agiu livremente e, por causa disso, não é moralmente responsável pelo que fez. Mas se este for o caso, parece que uma versão do Argumento da Manipulação é bem-sucedida. Considere o seguinte: Ernie tem um vizinho, Bert, um rapaz normal em todos os sentidos, muito similar a Ernie (com autocontrole ideal, racional, etc.), exceto pelo fato de que ele não foi criado por uma deusa (aspecto que, certamente, não é relevante para sua liberdade ou responsabilidade). Bert se encontra em circunstâncias muito similares àquelas que Ernie está, e Bert também julga, com base em sua deliberação racional, que o melhor a se fazer é **A**, e também age com base nesse julgamento, desse modo, fazendo com que **E** ocorresse. Não há diferença relevante entre Erine e Bert. Portanto, Bert também não agiu livremente e, também, não é moralmente responsável pelo que fez. Mas Bert, assim como Ernie, é normal em todo sentido, e podemos também estipular que ele, como Emie, satisfaz todas as condições compatibilistas plausíveis (históricas e ahistóricas) para ser um agente livre e moralmente responsável. Se ele age sem livre-arbítrio, então **todo** agente determinístico em **toda** ocasião age sem livre-arbítrio. Portanto, o tipo de liberdade necessária para a responsabilidade moral é incompatível com o determinismo.

Mele afirma que o caso de Ernie é melhor que os casos de manipulação anteriores de duas maneiras. Em primeiro lugar, as causas das ações de Ernie são

determinísticas, ao invés de causas que podem ser encontradas em mundos não-determinísticos. Além disso, é um caso em que claramente **não há diferença relevante** entre Ernie e qualquer caso de ação aparentemente livre e responsável em um mundo determinístico.

Mele está certo sobre o primeiro ponto. E, ainda que alguns tenham contestado isso (WALLER, 2014), deveríamos aceitar que ele também está certo sobre o segundo ponto.

Mas não deveríamos aceitar que o argumento seja bem-sucedido (deve-se notar que Mele não afirma que o é). Há um problema. Se realmente não há uma diferença relevante para a liberdade entre Bert e Ernie, por que deveríamos concluir, a partir da ausência de liberdade de Ernie, a ausência de liberdade de Bert ao invés do contrário, isto é, a partir da liberdade de Bert concluir a liberdade de Ernie?

No final das contas, nossa base para dizer que não há diferença relevante entre os dois são os **fatos históricos sobre a criação de Ernie**, que ele foi criado por uma deusa, com os poderes de um Demônio de Laplace, com certas intenções, e assim por diante, não são relevantes para a questão de saber se Ernie agiu livremente ou não 30 anos depois. Se não são relevantes, então não nos fornecem **razões** para pensar que Ernie não é livre. Por outro lado, temos motivos para pensar que Bert agiu livremente e é moralmente responsável pelo que fez; ele satisfaz as condições que usamos na vida real cotidiana, assim como todas as condições das melhores teorias compatibilistas que temos para oferecer. Se começamos com razões para pensar que Bert é livre e responsável (ou porque já somos compatibilistas ou porque nunca pensamos sobre o determinismo), o Argumento do Zigoto não dá qualquer razão para mudarmos de ideia. Para uma maior elaboração desta crítica, incluindo alguns contra-experimentos mentais úteis, veja Fischer (2011) e Kearns (2012).

Um(a) defensor(a) do Argumento do Zigoto pode responder afirmando que as intuições a favor da falta de liberdade e responsabilidade de Ernie são mais fortes que as intuições a favor da liberdade e responsabilidade de Bert. Mas isso é problemático. Intuições nem são sempre evidências confiáveis e, como já dissemos, há razões para ficarmos com um pé atrás em relação a nossas respostas intuitivas a descrições de mundos determinísticos: temos uma tendência natural de confundir "irá ser" com "deve ser". Há outras razões para não considerarmos como evidência a nossa intuição sobre a falta de liberdade de Ernie: ele foi criado por uma deusa, com o tipo de pré-ciência que nenhum humano poderia ter, com o único propósito de executar uma determinada ação 30 anos depois. Talvez nossas intuições sejam explicadas, mas não justificadas, pela crença de que ser criado desse modo tira de

Ernie a liberdade requerida para a responsabilidade. A primeira premissa do Argumento do Zigoto precisa ser defendida com algo além do apelo à intuição.

Para alguns trabalhos empíricos recentes sobre intuições sobre livre-arbítrio, determinismo e responsabilidade moral, consulte Nahmias *et al*. (2004), Nahmias (2011) e Murray e Nahmias (2014). Para algumas críticas do uso de intuições em Argumentos da Manipulação, veja Spitzley (2015).

Os(as) defensores(as) do Argumento do Zigoto não questionam esses pontos. O Argumento da Manipulação funciona apenas se a segunda premissa for verdadeira, e a segunda premissa diz que **não há diferenças relevantes** entre a Vítima, neste caso, Ernie, e qualquer caso determinístico normal de uma ação aparentemente livre e responsável, neste caso, Bert. Ernie difere de Bert em relação a certos fatos históricos sobre sua criação: o fato de que ele foi criado por uma deusa com presciência e intenções sobre seu futuro, então, nenhum desses fatos pode ser considerado relevante, **mesmo se eles afetarem nossas intuições**. Porém, um fato histórico é relevante segundo o(a) defensor(a) do argumento: o fato de que as causas determinísticas da ação de Ernie remontam a condições de 30 anos atrás, as quais não estão sob seu controle. A tese, então, é que Ernie agiu sem livre-arbítrio e sem responsabilidade porque o determinismo é verdadeiro. Mas esta tese era para ser a conclusão do argumento, não sua premissa.

4. Argumentos da Origem

> O que esse menino tem a ver com isso? Ele não era seu próprio pai; ele não era sua própria mãe; não era seus próprios avós. Tudo isso foi entregue a ele. Ele não se cercou de governantas e riqueza. Ele não se criou. E, ainda assim, ele é compelido a pagar. (DARROW, 1924, p. 65).

> Libertários e incompatibilistas não querem o indeterminismo por si só [...] o indeterminismo é uma dor de cabeça para eles, pois fica no meio do caminho e cria todo tipo de problema. O que eles querem é a responsabilidade última e a responsabilidade última requer o indeterminismo (KANE, 1989, p. 121).

Vejamos agora os argumentos a favor do incompatibilismo, baseados na ideia de que uma ação livre e responsável é causada e controlada por seu agente de um maneira especial em que "a cadeia acaba nele mesmo", incompatível com a causação entre eventos determinística.

O exemplo mais popular deste tipo de argumento é o qual chamaremos de "o argumento do advogado de defesa desesperado" (DARROW, 1924), que é bastante simples:

1. Meu cliente é responsável por seu crime somente se ele "criou a si próprio", isto é, somente se **ele** causou o fato de ser o tipo de pessoa que é.
2. Meu cliente não criou a si próprio.
3. Portanto, meu cliente não é responsável por seu crime.

O advogado de defesa tenta persuadir o júri de que seu cliente não é responsável por sua ação, mas não pelas condições que normalmente mostram que alguém não é moralmente responsável, como insanidade, acidentes, crenças erradas, coerção, deficiência mental, etc. Também não afirma que há algo que distingue seu cliente do resto de nós, mas seu argumento é que seu cliente não é responsável porque ele não criou a si próprio. Mas nenhum de nós cria a si mesmo (pelo menos não inteiramente), visto que somos todos produtos da hereditariedade e do ambiente. Logo, se aceitarmos o argumento do advogado de defesa, parece que estamos comprometidos com a conclusão de que ninguém nunca é moralmente responsável por algo.

Não é claro, entretanto, se temos aqui um argumento a favor do incompatibilismo. Se o temos, então há um argumento a favor do determinismo radical, isto é, um argumento a favor da tese de que o **determinismo** é verdadeiro e, **precisamente por isso**, não somos moralmente responsáveis.

Qual o argumento a favor da premissa 2? Afinal, nós criamos o nosso Eu em um certo sentido, naquele de que criamos outras coisas: plantamos flores, fazemos um jantar, construímos barcos, escrevemos livros e, ao longo das nossas vidas, reinventamo-nos, recriamo-nos e, de algum modo, "somos bem-sucedidos". É claro que não fazemos essas coisas "do nada", sem a ajuda de algo ou alguém, mas é impossível, ou, pelo menos, impossível para seres humanos, fazer qualquer coisa a partir do nada. A verdade ou a falsidade do determinismo não têm peso algum neste ponto.

Consulte G. Strawson (1986, 1994) sobre um argumento pela **impossibilidade** da "responsabilidade verdadeira", a qual se trata de uma versão mais sofisticada do argumento do advogado de defesa. Confira, também, Smilansky (2000). Para uma crítica do argumento de Strawson, veja Mele (1995) e Clarke (2005).

Se questionarmos nosso advogado de defesa, ou trouxermos uma filósofa ou filósofo para ajudá-lo, poderemos obter a seguinte resposta: esta ideia comum de que é possível, num mundo determinístico, para uma pessoa se criar, não é boa o bastante para o tipo de responsabilidade moral requerida para esta pessoa merecer culpa e punição.

Podemos admitir que nunca temos **controle completo** sobre as **ações** que executamos devido nossas escolhas (NAGEL, 1976), e isto **limita o controle** que temos sobre **nossa autocriação**. Mas somos moralmente responsáveis por nossas ações se tivermos **pelo menos algum controle** sobre nossa autocriação, e temos controle sobre nossa autocriação apenas se temos controle sobre nossas escolhas, que são as causas das ações por meio das quais nos criamos. E temos controle sobre nossas escolhas apenas se essas escolhas forem causadas por nós, **e nada nem ninguém nos levou** a promovê-las.

A premissa 2 é defendida pelo advogado de defesa por meio de sua interpretação da premissa 1, que a responsabilidade pelas nossas ações requer que "criemos a nós próprios" no sentido de que, ao longo das nossas vidas, fomos inúmeras vezes a **causa primeira** das escolhas que resultaram nas nossas ações, e, portanto, eventualmente levaram (embora de maneira que não pudéssemos controlar ou prever) a mudanças no nosso Eu. Desta maneira, o advogado mudou o foco do argumento; sai da exigência obviamente impossível de que a liberdade requerida para a responsabilidade moral consista **inteiramente** de um Eu autocriado, para a exigência (também intuitivamente plausível, mas não obviamente impossível) de que a liberdade requerida para a responsabilidade moral consista no que Robert Kane denominou "responsabilidade última", ou seja, que somos a fonte última ou a primeira causa de pelo menos algumas das nossas escolhas, causas de ações autoformadoras (KANE, 1996, 1999, 2008, 2011a). Confira, também, Chisholm (1964), Clarke (1993, 1996, 2003), O'Connor (1995a, 2000, 2011), Pereboom (2001, 2014) e Steward (2012).

Esta discussão nos leva à versão filosófica do argumento do advogado de defesa. Para algumas variações deste tipo de argumento, consulte Kane (1996, 1999, 2008, 2011a) e Pereboom (2001, 2005, 2014).

1. Agimos livremente (de modo necessário para termos responsabilidade moral) apenas se somos a origem última (originadores, a primeira causa) de algumas das nossas escolhas.
2. Se o determinismo for verdadeiro, então tudo o que fazemos é causado em última instância por eventos e circunstâncias fora do nosso controle.
3. Se tudo o que fazemos for causado em última instância por eventos e circunstâncias fora do nosso controle, então não somos a origem última (originadores, a primeira causa) das nossas ações.
4. Portanto, se o determinismo for verdadeiro, não somos a origem última das nossas ações.
5. Portanto, se o determinismo for verdadeiro, nunca agimos livremente e nunca somos moralmente responsáveis.

A premissa (2) se segue à definição de determinismo (se supusermos duas teses amplamente aceitas: que há causas em um universo determinístico e que a relação de causação é transitiva). Para questionamentos em relação ao último pressuposto, confira Hall (2000). A premissa (3) é claramente verdadeira. Logo, se quisermos rejeitar a conclusão, devemos rejeitar a premissa (1).

Compatibilistas argumentaram contra (1) de duas maneiras diferentes. Do lado positivo, defenderam que podemos dar uma explicação (de modo assumidamente esquivo) da noção de autodeterminação sem insistir que autodeterminação requer que sejamos as causas primeiras de nossas escolhas (FRANKFURT, 1971, 1988; WATSON, 1975, 1987, 2004; DENNETT, 1984; WOLF, 1990; WALLACE, 1994; FISCHER, 1994; FISCHER; RAVIZZA, 1998; BOK, 1998; NELKIN, 2011). Do lado negativo, compatibilistas atacaram (1) defendendo que (1) não tem serventia para o incompatibilismo: se aceitarmos (1), estamos comprometidos com a conclusão de que o livre-arbítrio e a responsabilidade moral são **impossíveis**, independentemente de o determinismo ser verdadeiro ou falso.

O desafio contra (1) toma a forma de um dilema: O determinismo é verdadeiro ou não. Se o determinismo for verdadeiro, então minhas escolhas são causadas em última instância por eventos e condições fora do nosso controle, então não sou a causa primeira das minhas escolhas; portanto, se aceitamos (1), segue-se que não sou livre nem responsável. Se o determinismo for falso, então algo que acontece em mim (algo que chamo de “minha escolha” ou “minha decisão”) pode ser o **primeiro evento** em uma cadeia causal que leva à sequência de movimentos que chamo de “minha ação”.

No entanto, já que o evento não é causalmente determinado, é apenas uma questão de probabilidade ou sorte saber se ele vai acontecer ou não; e esta questão não tem nada a ver **comigo**, não está **sob meu controle**, tal como o giro de uma roleta dentro do meu cérebro não está sob meu controle. Portanto, se o determinismo for falso, não sou a primeira causa ou a origem última das minhas escolhas, e, se aceitamos (1), não sou nem livre, nem responsável (AYER, 1954; WOLF, 1990).

Para defender (1) contra o dilema da "determinação ou aleatoriedade" mencionado anteriormente, o incompatibilista precisa oferecer uma explicação positiva da enigmática tese de que as **pessoas** são as primeiras causas de suas escolhas. A resposta incompatibilista tradicional é que esta tese precisa ser entendida literalmente, dizendo exatamente o que parece dizer. Nós – agentes, pessoas, coisas que perduram no tempo – somos causas com uma propriedade bastante especial: iniciamos cadeias causais, mas nada nem ninguém nos causa a assim o fazer. Como Deus, somos causadores não-causados, como um primeiro motor. Se Joe deliberadamente joga uma pedra que quebra uma janela, então, o evento da janela se quebrar foi causado pelo evento de Joe jogar a pedra, o qual foi causado pela escolha de Joe (outro evento). Porém, a escolha de Joe não foi causada por qualquer evento, nem mesmo pelo evento de Joe pensar que poderia ser legal jogar a pedra; a escolha foi causada pelo **próprio Joe**. E, dado que Joe não é um evento, ele não é o tipo de coisa que pode ser causada, ou pelo menos é isto que é sustentado por quem defende a causação do agente. A esse respeito, consulte Chisholm (1964), O'Connor (1995a, 2000, 2011) e Pereboom (2001).

Muitos filósofos e muitas filósofas pensam que a causação do agente é incoerente ou impossível, em virtude de certas considerações sobre a causação. Que sentido tem em dizer que uma pessoa, em oposição a uma **mudança** em uma pessoa, ou o **estado** de uma pessoa em um instante temporal, seja uma causa? (BOK, 1998). Sobre o assunto, confira, também, Clarke (2003) para um exame detalhado e simpático à metafísica da causação do agente, que termina com a conclusão de que há, no balanço geral, razões para se pensar que a causação do agente é impossível.

Outros (VAN INWAGEN, 2000; MELE, 2006) defenderam que, mesmo se a causação do agente for possível, ela não resolveria o problema de como transformar um evento indeterminado em um evento que está sob **nosso** controle, tal como nossas escolhas livres devem ser. E, ainda, outros argumentaram que, se a causação do agente é possível, ela o é em mundos determinísticos tanto quanto nos não-determinísticos (MARKOSIAN, 1999; FRANKLIN, 2016).

Uma diferente resposta incompatibilista ao dilema da "determinação ou aleatoriedade" apela à ideia de causação probabilística (KANE, 1996, 1999, 2008, 2011a). Se nossas escolhas são eventos que têm causas probabilísticas (por exemplo, nossas crenças, desejos e outras razões para agir), então, não parece mais plausível dizer que não temos controle sobre nossas escolhas. Fazemos escolhas por razões, e nossas razões causam nossas escolhas, ainda que indeterministicamente. Ora, a réplica de Kane pode ir na direção de evitar o segundo disjunto do dilema (a saber, a ausência de controle), mas não evita o primeiro disjunto. Se nossas escolhas são causadas por nossas razões, então nossas escolhas não são estarão entre as primeiras causas das nossas ações. E nossas razões são presumivelmente causadas, seja determinística ou probabilisticamente, de modo que a primeira causa das nossas ações não será uma razão. Mas então, em última instância, temos que nossas ações são causadas por eventos dos quais não temos controle algum, e portanto não somos a fonte última de nossas ações.

5. Escolha e o Argumento da Consequência

Pensamos que fazemos escolhas e que nossas escolhas normalmente fazem uma diferença no nosso futuro. Pensamos que faz sentido deliberar: como deliberamos, sobre quais razões consideramos, fazem uma diferença no que escolhemos e, portanto, no que fazemos. Também pensamos que, quando deliberamos, **realmente** há mais de uma escolha que somos capazes de fazer, mais de uma ação que podemos executar e mais de um futuro que, pelo menos parcialmente, está em nosso poder tornar realidade.

Nossas crenças sobre o que podemos fazer em relação ao futuro contrastam claramente com nossas crenças sobre nossa falta de poder em relação ao passado. Pensamos que **não** temos escolha **alguma** sobre o passado. Não deliberamos sobre o passado; pensamos que é irracional fazer algo com o objetivo de tentar mudar ou afetar o passado ("Você teve sua chance e vacilou. Agora já é tarde demais."). Nossas crenças sobre nossas opções, oportunidades, alternativas, possibilidades, habilidades, poderes, e assim por diante, são todas direcionadas ao futuro. Podemos resumir esse contraste dizendo que pensamos que o futuro é "aberto" em algum sentido que contrasta com a não-abertura ou "fixidez" do passado.

Embora não pensemos que, no agora, temos uma escolha sobre o passado, temos crenças sobre o que **era** possível a nós em um passado. Quando tentamos

explicar por que fizemos o que fizemos, ou quando nos culpamos e nos reprovamos, ou simplesmente consideramos se fizemos o certo (ou o mais sensato, racional e assim por diante), avaliamos nossa ação ao compará-la com nossa crença sobre quais eram nossas outras ações possíveis **naquele momento**. Apenas nos culpamos, criticamos, censuramos ou nos arrependemos, etc., quando acreditamos que tínhamos alternativas. E se posteriormente descobrirmos que estávamos enganados em crer que uma ação **X** era uma dessas alternativas, pensamos que é irracional nos criticarmos ou nos arrependermos por não ter executado **X**.

O determinismo é compatível com a verdade dessas crenças? Em particular, é compatível com a crença de que, frequentemente, somos capazes de escolher e executar mais de uma ação?

A resposta incompatibilista tradicional é "Não", e é fácil ver o porquê. Se por "pode" queremos dizer algo no sentido sugerido pela visão padrão do "futuro em aberto", então é tentador pensar que o passado é necessário num sentido absoluto. E é natural pensarmos que poderíamos agir de outro modo apenas se pudéssemos agir de outro modo dado o passado; isto é, apenas se o nosso agir de outro modo for uma continuação possível do passado. Se seguirmos essa linha de raciocínio, concluiremos que podemos agir de outro modo apenas se o nosso agir de outro modo for uma continuação possível do passado e consistente com as leis da natureza. Mas se o determinismo for verdadeiro, há apenas uma continuação possível do passado que seja consistente com as leis da natureza. E assim chegamos à conclusão incompatibilista, nosso futuro, **de fato**, é o único futuro possível. O que de fato fazemos é a única coisa que podemos fazer.

No entanto, o argumento acima é demasiado apressado. Há uma explicação alternativa para nossas crenças sobre o futuro em "aberto" em oposição ao passado "fixo" – a direção da causação. Cadeias causais vão do passado para o futuro, não na direção oposta. Nossas deliberações causam nossas escolhas, as quais causam nossas ações; mas não o oposto. Nossas escolhas causam eventos futuros; elas nunca causam eventos passados. Saber exatamente por que a causação funciona desse modo é uma questão profunda e difícil, mas a posição mais aceita, entre filósofos e filósofas da ciência, é que a assimetria temporal da causação é um fato fundamental, porém contingente, sobre o nosso universo. Se as coisas fossem diferentes o bastante, ou seja, se pudéssemos viajar no tempo de volta ao passado, então teríamos a capacidade que, de fato, não temos, de causalmente afetar as pessoas e as coisas do passado (HORWICH, 1987; D. LEWIS, 1976). Se assim o for, então não precisamos supor que o passado seja metafisicamente ou absolutamente

necessário para explicar o contraste entre o passado fixo e o futuro aberto. O passado **poderia ter sido diferente**. Mas, em virtude de como as coisas de fato são, dado o fato contingente de que a causação vai do passado para o futuro, não **somos capazes de fazer algo que fizesse o passado ser diferente**.

Essa explicação alternativa da nossa crença comum sobre o contraste entre o passado fixo e o futuro em aberto permite à(ao) compatibilista dizer o tipo de coisa que compatibilistas tradicionalmente querem dizer: O "pode", do livre-arbítrio e da liberdade de ação, é o "pode" da dependência causal e contrafactual. Nosso futuro é aberto porque depende, causalmente e contrafactualmente, das nossas escolhas, as quais por sua vez dependem, causalmente e contrafactualmente, da nossa deliberação e das razões que julgamos ter. Pelo menos no caso normal, no qual não há limitação externa, nem compulsão interna ou alguma patologia. Se nossas razões fossem diferentes, de um modo apropriado, teríamos escolhido de outro modo e, se tivéssemos escolhido de outro modo, teríamos agido de outro modo. E nossas razões **podem** ser diferentes, pelo menos no sentido em que nós, diferentemente de criaturas mais simples e as crianças pequenas, temos a capacidade de avaliar criticamente nossas razões, crenças, desejos, valores, princípios, etc., e que temos, e, às vezes, exercitamos, **a capacidade de mudar** nossas razões (BOK, 1998; DENNETT, 1984; FISCHER, 1994; FISCHER; RAVIZZA, 1998; FRANKFURT, 1971, 1988; LEHRER, 1976, 1980, 2004; WALLACE, 1994; WATSON, 1975, 1987, 2004; WOLF, 1990; SMITH, 1997, 2004; PETTIT; SMITH, 1996; NELKIN, 2011). Tudo isso é compatível com o determinismo. Portanto, a verdade do determinismo é compatível com a verdade da nossa crença de senso comum de que realmente temos uma escolha sobre o futuro e que realmente podemos escolher e agir de um modo diferente do que escolhemos e agimos de fato.

Incompatibilistas pensam que esta e **qualquer** explicação compatibilista do "pode" na liberdade do arbítrio está, e *deve* estar, errada. O Argumento da Consequência (GINET, 1966, 1983, 1990; VAN INWAGEN, 1975, 1983, 2000; WIGGINS, 1973; LAMB, 1977) é amplamente considerado o melhor argumento por esta conclusão. No restante desta seção, vamos investigar mais atentamente a versão de van Inwagen deste influente e importante argumento.

Em *An Essay on Free Will* (1983), van Inwagen apresenta três argumentos formais que, segundo ele, são três versões do mesmo argumento básico, que o autor caracteriza da seguinte maneira:

> Se o determinismo for verdadeiro, então nossas ações são a consequência das leis da natureza e dos eventos no passado remoto. Mas não temos escolha sobre o que aconteceu antes de nascermos, nem temos escolha sobre quais são as leis da natureza. Portanto, não temos escolha sobre a consequência dessas coisas (incluindo nossas ações presentes).

Comecemos com a terceira versão do argumento, o argumento da Regra Beta. Esse argumento usa um operador sentencial modal que van Inwagen define da seguinte maneira: **'N*p*'** abrevia **'*p* e ninguém tem, nem nunca teve, escolha sobre *p*'**. Van Inwagen nos diz que a lógica de **'N'** inclui estas duas regras de inferência, em que □***p*** assere que isso é logicamente necessário que ***p***:

Alfa: De □***p***, podemos inferir **N*p***.

Beta: De **N*p*** e **N(*p* ⊃ *q*)**, podemos inferir **N*q***.

No argumento que se segue, **'*L*'** é uma abreviação de uma sentença expressando a conjunção de todas as leis da natureza, **'*H*'** é a sentença expressando a proposição verdadeira sobre o estado total do mundo em um instante temporal no passado distante, antes de qualquer agente existir, já **'□'** é "é logicamente necessário que". Por sua vez, **'⊃'** é a implicação material e **'*P*'** qualquer sentença que exprima uma proposição verdadeira.

O argumento é uma prova condicional: assumimos o determinismo e mostramos que disso se segue que ninguém tem, nem nunca teve, escolha sobre **qualquer** proposição verdadeira, incluindo proposições sobre ações aparentemente livres de seres humanos.

1	□ ((***H*** & ***L***) ⊃ ***P***)	definição de determinismo;
2	□ (***H*** ⊃ (***L*** ⊃ ***P***))	de 1 por lógica modal proposicional;
3	**N**(***H*** ⊃ (***L*** ⊃ ***P***))	de 2 por regra Alfa;
4	**N*H***	premissa, fixidez do passado;
5	**N**(***L*** ⊃ ***P***)	de 3 e 4 por regra Beta;
6	**N*L***	premissa, fixidez das leis;
7	**N*P***	de 5 e 6 por regra Beta;

As premissas (1) e (2) se seguem do determinismo. (3) segue de (2) pela aplicação da regra Alfa. A regra Alfa parece incontroversa, mas confira o que diz Spencer (2017).

As premissas (4) e (6) também parecem incontroversas. A necessidade de ***N*** não é a necessidade lógica ou metafísica. Podemos insistir que as leis e o passado distante *poderiam*, no sentido metafísico, ter sido diferentes, então nem □***H*** nem □***L*** são verdadeiros. Ainda assim, parece inegavelmente verdadeiro que **não temos escolha sobre** as leis e o passado distante serem como são; não somos capazes de fazer algo que faria com que fosse o caso que as leis ou o passado distante fossem diferentes do que realmente são.

A regra Beta é a chave do argumento. É o que faz com que esta versão do Argumento da Consequência não seja um argumento amplamente aceito como falacioso.

□ (***P*** ⊃ ***Q***)
P
Portanto, □***Q***

Um exemplo desta inferência inválida é o argumento, muitas vezes, denominado por "falácia fatalista":

□ (é verdade que irá chover amanhã ⊃ irá chover amanhã);
É verdade que irá chover amanhã;
Portanto, □ (irá chover amanhã);

Outro exemplo:

□ ((***H*** & ***L***) ⊃ ***P***)
H & ***L***
Portanto, □***P***

Por outro lado, a seguinte inferência é válida:

□***P***
□ (***P*** ⊃ ***Q***)
Portanto, □***Q***

A necessidade expressa pelo operador de “não ter escolha sobre” não é a necessidade metafísica ou lógica. Ainda assim, parece similar o bastante para Beta ser uma regra de inferência válida. Ou, pelo menos, foi o que van Inwagen defendeu por meio de exemplos:

N (O sol explode no ano 2000);
N (O sol explode no ano 2000 ⊃ Toda a vida na Terra acaba no ano 2000);
Portanto, **N** (Toda a vida na Terra acaba no ano 2000);

Uma resposta antiga ao Argumento da Consequência consistia em defender que Beta é inválida porque uma teoria compatibilista da habilidade de agir de outro modo está correta (GALLOIS, 1977; FOLEY, 1979; SIOTE, 1982; FLINT, 1987). Por exemplo, se ***“S*** é capaz de fazer ***X”*** significar “Se ***S*** tentasse fazer ***X***, ***S*** faria ***X*”**, então, as premissas do argumento são verdadeiras (já que, mesmo se ***S*** tentasse mudar as leis ou o passado, ***S*** não conseguiria), mas a conclusão é falsa (já que o determinismo é consistente com a verdade de contrafactuais como “se ***S*** tentasse levantar a mão, ***S*** a levantaria”).

Incompatibilistas não se convenceram com esta resposta, dizendo que, na realidade, a validade de Beta é mais plausível que a verdade de qualquer teoria compatibilista sobre a habilidade de agir de outro modo. Apontam que não há concordância, mesmo entre compatibilistas, sobre qual deveria ser essa teoria, e as teorias mais simples (as chamadas “Análises Condicionais”, originalmente propostas por Hume) foram rejeitadas, mesmo por compatibilistas.

Para críticas às Análises Condicionais, consulte Austin (1961), Lehrer (1968, 1976) e van Inwagen (1983). Para defesas de uma teoria compatibilista da habilidade de agir de outro modo, veja Moore (1912), Hobart (1934), Kapitan (1991, 1996, 2011), Lehrer (1980, 2004), Bok (1998), Smith (1997, 2004), Campbell (2005), Perry (2004), Vihvelin (2004, 2013) e Fara (2008).

Mais recentemente, van Inwagen (2000) aceitou que Beta é inválida. McKay e Johnson (1996) mostraram que Beta implica Aglomeração:

N*p*
N*q*
Portanto, **N**(***p*** & ***q***)

Aglomeração é certamente inválida. Para ver isso, seja ***'p'*** a abreviação de "na moeda não sai cara", e seja ***'q'*** a abreviação de "na moeda não sai coroa", suponha que a moeda seja comum e que não tenha sido arremessada, mas que alguém poderia tê-la arremessado (MCKAY; JOHNSON, 1996).

Para contraexemplos à Beta, confira Widerker (1987), Huemer (2000) e Carlson (2000).

Van Inwagen propôs consertar o Argumento da Consequência substituindo **'N'** por **'N*'**, em que **'N****p***'** significa "***p*** e ninguém pode, nem nunca pôde, fazer algo tal que, se fizesse, ***p*** poderia ser falso". Aglomeração é válida para '**N***', e, portanto, essa objeção em particular à validade de Beta não se aplica.

Também se sugeriu (FINCH; WARFIELD 1998) que o Argumento da Consequência poderia ser consertado mantendo-se '**N**' e substituindo Beta por Beta 2:

Beta 2 De **N*p*** e □ (***p*** ⊃ ***q***), podemos inferir **N*q***

Com isto, teríamos o seguinte argumento:

1	**N** (***L*** & ***H***)	fixidez das leis e do passado;
2	□ ((***L*** & ***H***) ⊃ ***P***)	determinismo;
3	**N*P***	de 1 e 2 por Beta 2

Outros modos de consertar o argumento foram propostos por O'Connor (1993) e Huemer (2000).

Então ainda parece que o compatibilismo está com problemas. Porque parece plausível supor que não somos capazes de fazer algo que poderia falsificar *H* ou *L*. E parece plausível supor que nós não temos escolha sobre se (*H*&*L*). Precisamos portanto ir mais fundo para criticar esse argumento.

David Lewis (1981) nos diz para conceber o argumento como uma redução ao absurdo. Um compatibilista é alguém que defende que a verdade do determinismo é compatível com a existência dos tipos de capacidades que assumimos termos em situações típicas de escolha. Vamos chamá-las de "capacidades comuns". O Argumento da Consequência, como Lewis o articula, diz que, se assumirmos que um agente determinístico tem as capacidades ordinárias, somos forçados a concedê-lo também capacidades inacreditáveis.

Vejamos, com algumas pequenas modificações, a formulação de Lewis do argumento:

> Suponhamos que o determinismo seja verdadeiro, e que não levantei minha mão (naquela reunião de departamento para votar naquela proposta) mas tinha a capacidade para levantá-la. Se eu tivesse exercido minha capacidade – se tivesse levantado minha mão – então ou o passado remoto ou as leis da natureza teriam sido diferentes (teriam que ter sido diferentes). Mas se isso é o caso, então tenho ao menos uma de duas capacidades inacreditáveis – a capacidade de mudar o passado remoto ou a capacidade de mudar as leis. Mas supor que tenho qualquer uma destas habilidades é absurdo. Então devemos rejeitar a afirmação de que eu tinha a capacidade de levantar minha mão.

Esta versão contrafactual do Argumento da Consequência destaca bem um ponto que a versão da regra Beta encobre. O argumento depende de uma tese sobre **contrafactuais**. O argumento diz que, se o determinismo for verdadeiro, então, pelo menos um destes contrafactuais é verdadeiro:

> **Passado Diferente:** Se tivesse levantado minha mão, o passado remoto teria sido diferente.
> **Leis Diferentes:** Se tivesse levantado minha mão, as leis teriam sido diferentes.

Ambos contrafactuais parecem inacreditáveis para muitas pessoas. Mas há uma razão para isto: não estamos acostumados a pensar nos termos do determinismo e não estamos acostumados a especular contrafactualmente sobre **o que teria sido o caso anteriormente**, se alguma coisa em um mundo determinístico tivesse acontecido de forma que não do modo como de fato aconteceu.

Por outro lado, somos bons em avaliar contrafactuais, ou pelo menos alguns contrafactuais, e somos especialmente bons em avaliar aqueles contrafactuais que consideramos em contextos de escolha, quando nos perguntamos sobre as consequências causais de nossas ações contempladas (O que aconteceria se... Eu acendesse este fósforo, colocasse meu dedo no fogo, jogasse esta pedra naquela janela, levantasse minha mão?). E quando contemplamos nossas opções, tomamos como dado a existência de muitos fatos, incluindo fatos sobre as leis e o passado.

Em outras palavras, quando avaliamos contrafactuais na vida real, fazemos isso considerando situações imaginárias que são muito parecidas com a situação em que de fato estamos e não supomos existir qualquer dissimilaridade **gratuita** daquilo que é, de fato, o caso. E supor uma diferença no passado ou nas leis parece uma diferença gratuita.

Então não é uma surpresa que quando nossa atenção é direcionada para **Passado Diferente** *e* **Leis Diferentes**, esses contrafactuais nos parecem inacreditáveis ou, pelo menos, **esquisitos**. Mas isso não significa que eles sejam falsos e, se o determinismo for verdadeiro, então ou **Passado Diferente** ou **Leis Diferentes** *é* verdadeiro.

Então, o primeiro ponto é que todos precisamos de uma teoria de contrafactuais e, se o determinismo for verdadeiro, as contrafactuais verdadeiras incluirão ou **Passado Diferente** ou **Leis Diferentes**.

O segundo ponto é que os detalhes da solução compatibilista correta ao problema do livre-arbítrio/determinismo dependerão dos detalhes da teoria correta de contrafactuais.

Para críticas similares ao Argumento da Consequência, consulte Fischer (1983, 1988), Horgan (1985), Watson (1987) e Vihvelin (1988).

Se a teoria de contrafactuais de Lewis (1973, 1979) está correta, ou mais ou menos correta, então, os contrafactuais relevantes sobre o passado e as leis, em um mundo determinístico, são:

> **Praticamente o Mesmo Passado:** Se eu tivesse levantado minha mão, o passado teria sido exatamente o mesmo até um instante temporal um pouco antes do momento da minha decisão de levantar minha mão.
>
> **Leis Levemente Diferentes:** Se eu tivesse levantado minha mão, as leis teriam sido levemente diferentes de modo a permitir uma divergência do curso correto da história atual um pouco antes do momento da minha decisão de levantar minha mão.

Por outro lado, se a teoria de Lewis estiver errada, e contrafactuais sempre forem avaliados mantendo as leis fixas, então as contrafactuais relevantes, em um mundo determinístico, são:

> **Mesmas Leis** Se tivesse levantado minha mão, as leis ainda teriam sido exatamente as mesmas.

> **Passado Diferente** Se tivesse levantado minha mão, a história passada teria sido diferente desde o Big Bang.

Para críticas à teoria de Lewis e defesas da teoria de contrafactuais que mantém as leis fixas, veja Bennett (1984) e Dorr (2016).

Precisamos fazer uma escolha. Todos precisamos de uma teoria de contrafactuais, e nossa teoria deve fornecer os vereditos corretos para as contrafactuais certamente verdadeiras em mundos determinísticos assim como em indeterminísticos. Qual teoria escolhemos não tem a ver com o problema livre-arbítrio/determinismo e tudo a ver com o modo como avaliamos as contrafactuais.

Podemos explicar agora a essência da resposta de Lewis à versão contrafactual do Argumento da Consequência de um modo que não requer que aceitemos a teoria de contrafactuais de Lewis.

O argumento se vale de uma ambiguidade entre dois contrafactuais.

> (C1) Se eu tivesse levantado minha mão, as leis (ou o passado) teria sido diferente.
> (C2) Se eu tivesse levantado minha mão, eu teria, por causa disso, **causado** as leis (ou o passado) serem diferentes.

Há uma ambiguidade correspondente entre as duas afirmações sobre capacidades:

> (A1) Eu tinha a capacidade de fazer algo (levantar minha mão) tal que, se tivesse exercido essa capacidade, as leis (ou o passado) teriam sido diferentes.
> (A2) Eu tinha a capacidade de fazer algo (levantar minha mão) tal que, seu tivesse exercido essa capacidade, eu teria, por causa disso, **causado** as leis (ou o passado) serem diferentes.

O problema com o argumento, diz Lewis, é que ele é ambíguo entre essas duas afirmações sobre capacidades. Para contar como uma redução ao absurdo do compatibilismo, o argumento precisa estabelecer que o compatibilismo está comprometido com (A2). Mas o compatibilismo está comprometido apenas com (C1) e, portanto, apenas com (A1). O compatibilismo está comprometido apenas com o fato que, se o determinismo for verdadeiro, temos capacidades que exercitaríamos **apenas** *se* o passado (ou as leis) tivessem sido diferentes do modo apropriado. E, ainda que isto

possa soar estranho, não é mais inacreditável do que a tese de que o exercício exitoso de nossas capacidades depende não apenas de nós, mas também da cooperação de coisas fora do nosso controle: a sorte ou azar de nossos arredores imediatos. Dado que não somos nem super-heróis nem deuses, sempre estamos nesta posição, independentemente da verdade ou falsidade do determinismo.

O Argumento da Consequência foi proposto com a intenção de ser um argumento que partia de premissas que todos precisamos aceitar, premissas sobre nossa falta de controle sobre o passado e as leis, e concluía que, se o determinismo for verdadeiro, não temos o livre-arbítrio que o bom senso diz possuirmos. A versão contrafactual do argumento afirma que se atribuímos capacidades comuns a agentes determinísticos, somos forçados a concedê-lo também capacidades inacreditáveis de mudar o passado ou as leis. Mas tais conclusões inacreditáveis não se seguem. Tudo que se segue é algo que precisamos aceitar de qualquer modo, como um preço de nossa natureza não-divina: que o exercício de nossas capacidades sempre depende, em parte, de circunstâncias fora de nosso controle. A esse respeito, confira, também, Fischer (1983, 1988, 1994), Horgan (1985), Vihvelin (2011, 2013), Kapitan (1991, 1996, 2011), Carlson (2000) e Schneider (2004).

Se o objetivo do Argumento da Consequência era mostrar que nenhuma teoria compatibilista sobre "poderia ter agido de outro modo" **pode** funcionar, então, Lewis, certamente, está correto, a redução ao absurdo falha. A distinção entre (A1) e (A2) permite ao compatibilista evitar fazer afirmações inacreditáveis sobre nossas capacidades como agentes determinísticos livres. Por outro lado, o incompatibilista certamente tem um ponto quando afirma que é difícil acreditar que alguém tem a capacidade descrita em (A1). Acreditamos que nossas capacidades como agentes são limitadas pelo passado e pelas leis. Um jeito de entender como esta crença é compatível com o determinismo é o seguinte: **não temos poder causal sobre o passado e as leis**. Mas é natural entender a limitação de um modo diferente, mais simples: somos capazes de fazer apenas aquelas coisas de modo tal que, ao fazê-las, não se requer contrafactualmente uma diferença no passado ou nas leis. E isto leva mais ou menos diretamente à conclusão incompatibilista de que, se o determinismo for verdadeiro, então nunca somos capazes de agir de outro modo.

Isto nos leva de volta ao nosso ponto de partida. Nossa rede de crenças de senso-comum sobre nós mesmos como deliberadores, escolhedores e agentes inclui a crença de que o futuro é aberto em um sentido em que o passado não é. Ela também inclui a crença de que nossas capacidades e poderes são limitados pelas leis. Um jeito de entender estas crenças leva ao incompatibilismo, outro jeito

não. Qual deles está correto?

O Argumento da Consequência é uma tentativa de fornecer um **argumento** em defesa do modo incompatibilista de entender estas crenças de senso-comum. Mesmo se ele falhar como uma redução ao absurdo, obteve sucesso de outros modos, pois tornou claro que o problema do livre-arbítrio/determinismo é um problema metafísico e que os detalhes de sua solução envolvem questões sobre nossas capacidades e poderes, assim como questões mais gerais sobre a natureza da causação, contrafactuais e leis da natureza. As capacidades ou poderes de escolhedores e agentes podem ser entendidos como um tipo de capacidade natural ou disposição? Há uma alternativa incompatibilista viável? Como deveríamos entender as contrafactuais sobre ações e escolhas alternativas de agentes em mundos determinísticos? A proposta compatibilista sobre o modo em que as leis e o passado nos limitam é defensável? Os incompatibilistas estão comprometidos com uma visão particular sobre a natureza das leis da natureza? Eles estão comprometidos coma rejeição de uma visão Humiana, por exemplo?

Na medida em que o Argumento da Consequência nos direcionou para estas profundas e difíceis questões metafísicas subjacentes, ele representa um passo significativo na discussão de um dos mais intratáveis problemas da filosofia. Para discussões sobre algumas destas questões, consulte D. Lewis (1973, 1979, 1997), Smith (1997, 2004), Pettit Smith (1996), C. Taylor e Dennett (2002), Berofsky (2003, 2012), Beebee e Mele (2002), Beebee (2003), Oakley (2006), Graham (2008), Pendergraft (2011), Vihvelin (2004, 2013, 2017), Clarke (2009), Perry (2004, 2008), van Inwagen (2004a), Fara (2008), Mackie (2003, 2014b), Schlosser (no prelo) e Spencer (2017).

Referências

ADAMS, R. M. Theories of Actuality. **Noûs**, v. 8, n. 3, p. 211-231, 1974. DOI:10.2307/2214751

ALBERT, D. Z. **Quantum Mechanics and Experience**. Cambridge: Harvard University Pres, 1992.

ALBRITTON, R. Freedom of Will and Freedom of Action. *In:* WATSON, G. (ed.). **Free Will**. 2. ed. Oxford: Oxford University Press, 2003.

ALVAREZ, M. Actions, Thought Experiments, and the 'Principle of Alternate Possibilities'. **Australasian Journal of Philosophy**, v. 85, n. 1, p. 61-81,

2009. DOI:10.1080/00048400802215505

ANSCOMBE, G. E. M. **"Causality and Determination"**. Minneapolis: University of Minnesota Press,1981, p. 133-147.

ARMSTRONG, D. **What is a Law of Nature?** Cambridge: Cambridge University Press, 1983.

AUSTIN, J. L. Ifs and Cans. **Proceedings of the British Academy**, v. 42, p. 109-132, 1956. DOI:10.1093/019283021X.003.0009

AYER, A.J. Freedom and Necessity. *In:* PEREBOOM, D. (ed.). **Free Will**. *2. ed.* Indianapolis: Hackett Publishing, 2009a.

BAKER, L. R. The Irrelevance of the Consequence Argument. **Analysis**, v. 68, n. 1, p. 13-22, 2008. DOI: 10.1093/analys/68.1.13.

BAILEY, A. M. Incompatibilism and the Past. **Philosophy and Phenomenological Research**, v. 85, n. 2, p. 351-376, 2012. DOI: 10.1111/j.1933-1592.2011.00542.x.

BALAGUER, M. **Free Will as an Open Scientific Problem**. Cambridge: MIT Press, 2010.

BEEBEE, H. The Non-Governing Conception of Laws. **Philosophy and Phenomenological Research**, v. 61, n. 3, p. 571-594, 2000. DOI: 10.2307/2653613.

BEEBEE, H. Reply to Huemer on the Consequence Argument. **The Philosophical Review**, v. 111, n. 2, p. 235-241, 2002. DOI: 10.1215/00318108-111-2-235.

BEEBEE, H. Local Miracle Compatibilism. **Noûs**, v. 37, n. 2, p. 258-277, 2003. DOI: 10.1111/1468-0068.00438.

BEEBEE, H. **Free Will:** an introduction. New York: Palgrave Macmillan, 2013.

BEEBEE, H.; MELE, A. Humean Compatibilism. **Mind**, v. 111, n. 442, p. 201-23, 2002. DOI: 10.1093/mind/111.442.201.

BENNETT, J. Counterfactuals and Temporal Direction. **The Philosophical Review**, v. 93, n. 1, p. 57-91, 1984. DOI: 10.2307/2184413.

BENNETT, J. **A Philosophical Guide to Conditionals**. Oxford: Clarendon Press, 2003. DOI: 10.1093/0199258872.001.0001.

BEROFSKY, B. (ed.). **Free will and determinism.** New York: Harper & Row, 1966. BEROFSKY, B. The Irrelevance of Morality to Freedom. **Bowling Green Studies in Applied Philosophy**, v. 2, p. 38-47, 1980.

BEROFSKY, B. Classical Compatibilism: Not Dead Yet. *In:* WIDERKER, D.; MCKENNA, M. (ed.). **Moral responsibility and alternative possibilities**. Burlington: Ashgate Publishing Company, 2003, p. 107-126.

BEROFSKY, B. The Myth of Source. **Acta Analytica**, v. 21, n. 4, p. 3-18, 2006a.

BEROFSKY, B. Global Control and Freedom. **Philosophical Studies**, v. 131, n. 2, p. 419-445, 2006b. DOI:10.1007/s11098-004-7490-1

BEROFSKY, B. **Nature's Challenge to Free Will**. Oxford: Oxford University Press, 2012. DOI: 10.1093/acprof:oso/9780199640010.001.0001.

BOK, H. **Freedom and responsibility**. Princeton: Princeton University Press, 1998.

BRAND, M.; WALTON, D. (ed). **Action theory:** Proceedings of the Winnipeg Conference on Human Action, held at Winnipeg. Dordrecht: Reidel,1976.

CAMPBELL, J. K. Compatibilist Alternatives. **Canadian Journal of Philosophy**, v. 35, n. 3, p. 387-406, 2005. DOI: 10.1080/00455091.2005.10716595.

CAMPBELL, J. K. Free Will and the Necessity of the Past. **Analysis**, v. 67, n. 2, p. 105-111, 2007. DOI: 10.1093/analys/67.2.105.

CAMPBELL, J. K. The Consequence Argument. *In:* GRIFFITH, M.; LEVY, N.; TIMPE, K. **Routledge Companion to Free Will**. New York: Routledge, 2017.

CAMPBELL, J. K.; O'ROURKE, M.; SHIER, D. (ed.). **Freedom and Determinism**. Press: Bradford Books, 2004.

CAMPBELL, J. K.; O'ROURKE, M.; SLATER, M. H. (ed.). **Carving Nature at its Joints:** Natural Kinds in Metaphysics and Science. Cambridge: MIT Press, 2011. DOI: 10.7551/mitpress/9780262015936.001.0001.

CAPES, J. A. Action, Responsibility, and Ability to Do Otherwise. **Philosophical Studies**, v. 158, n. 1, p. 1-15, 2012. DOI: 10.1007/s11098-010-9662-5.

CARLSON, E. Incompatibilism and the Transfer of Power Necessity. **Noûs**, v. 34, n. 2, p. 277-290, 2000. DOI: 10.1111/0029-4624.00211.

CARLSON, E. On a New Argument for Incompatibilism. **Philosophia**, v. 31, n. 1–2, p. 159-164, 2003. DOI: 10.1007/BF02380930.

CARROLL, J. Nailed to Hume's Cross? *In:* SIDER, T.; HAWTHORNE, J.; ZIMMERMAN, D. W. (ed.). **Contemporary Debates in Metaphysics**. Malden: Blackwell Publishing, 2008, p. 67-81.

CHISHOLM, Roderick M. Human Freedom and the Self. *In:* Watson, G. (ed.). **Free Will**. 2. ed. Oxford: Oxford University Press, 2003.

CLARKE, R. Towards a Credible Agent-Causal Account of Free Will. **Noûs**, v. 27, n. 2, p. 191-203. DOI: 10.2307/2215755.

CLARKE, R. Agent Causation and Event Causation in the Production of Free Action. **Philosophical Topics**, v. 24, n. 2, p. 19-48, 1996. DOI: 10.5840/philtopics19962427.

CLARKE, R. **Libertarian Accounts of Free Will**. New York: Oxford University Press, 2003. DOI: 10.1093/019515987X.001.0001.

CLARKE, R. On an Argument for the Impossibility of Moral Responsibility. **Midwest Studies in Philosophy**, v. 29, p. 13-24, 2005. DOI: 10.1111/j.1475-4975.2005.00103.x.

CLARKE, R. Dispositions, Abilities to Act, and Free Will: The New Dispositionalism, **Mind**, v. 118, n. 470, p. 323-351, 2009. DOI: 10.1093/mind/fzp034.

CULLITY, G.; GAUT, B. (ed.). **Ethics and Practical Reason**. New York: Clarendon Press, 1997.

DARROW, C. The Trial of Leopold and Loeb, Attorney for the Damned: Clarence Darrow. *In:* WEINBERG, A. (ed.) *T***he Courtroom.** Chicago: The University of Chicago Press, 1924, p. 16-86.

DEERY, O. The Fall from Eden: Why Libertarianism Isn't Justified by Experience. **Australasian Journal of Philosophy**, v. 93, n. 2, p. 319-334, 2015. DOI: 10.1080/00048402.2014.968596.

DEERY, O. Why People Believe in Indeterminist Free Will. **Philosophical Studies**, v. 172, n. 8, p. 2033-2054, 2015. DOI: 10.1007/s11098-014-0396-7.

DEERY, O.; BEDKE, M.; NICHOLS, S. Phenomenal Abilities: Incompatibilism and the Experience of Agency. *In:* SHOEMAKER, D. **Oxford Studies in Agency and responsibility**, 2013, p. 126-150. DOI: 10.1093/acprof:oso/9780199694853.003.0006.

DENNETT, D. **Elbow Room:** the varieties of free will worth wanting. Cambridge: MIT Press, 1984.

DORR, C. Against Counterfactual Miracles. **The Philosophical Review**, v. 125, n. 2, p. 241-286, 2016. DOI: 10.1215/00318108-3453187.

DOUBLE, R. Puppeteers, Hypnotists, and Neurosurgeons. **Philosophical Studies**, v. 56, n. 2, p. 163-173, 1989. DOI: 10.1007/BF00355940.

DOUBLE, R. **The Non-Reality of Free Will**. New York: Oxford University Press, 1991.

EARMAN, J. **A Primer on Determinism**. Dordrecht: D. Reidel Publishing Company, 1986.

EARMAN, J. Determinism: What We Have Learned and What We Still Don't Know. In: CAMPBELL, J. K.; O'ROURKE, M.; SHIER, D. (ed.). **Freedom and Determinism**. Press: Bradford Books, 2004, p. 21-46.

EKSTROM, L. **Free Will**: A Philosophical Study. Boulder: Westview Press, 2000.

FARA, M. Masked Abilities and Compatibilism, **Mind**, v. 117, n. 468, p. 843-865, 2008. DOI:10.1093/mind/fzn078

FELDMAN, R. Freedom and Contextualism. *In:* CAMPBELL, J. K.; O'ROURKE, M.; SHIER, D. (ed.). **Freedom and Determinism**. Press: Bradford Books, 2004, p. 255-276.

FINCH, A. On Behalf of the Consequence Argument: Time, Modality, and the Nature of Free Action. **Philosophical Studies**, v. 163, n. 1, p. 151-170, 2013. DOI: 10.1007/s11098-011-9791-5.

FINCH, A.; WARFIELD, T. The *MIND* Argument and Libertarianism. **Mind**, v. 107, n. 427, p. 515-528, 1998. DOI: 10.1093/mind/107.427.515.

FISCHER, J. M. Incompatibilism. **Philosophical Studies**, v. 43, n. 1, p. 121-137, 1983. DOI: 10.1007/BF01112527.

FISCHER, J. M. Power Necessity. **Philosophical Topics**, v. 14, n. 2, p. 77-91, 1986. DOI: 10.5840/philtopics19861424.

FISCHER, J. M. Freedom and Miracles, **Noûs**, v. 22, n. 2, p. 235-252, 1988. DOI: 10.2307/2215861.

FISCHER, J. M. **The Metaphysics of Free Will:** An Essay on Control. Oxford: Blackwell, 1994.

FISCHER, J. M. Frankfurt-Style Compatibilism. *In:* WATSON, G. (ed.). **Free Will**. 2. ed. Oxford: Oxford University Press, 2003.

FISCHER, J. M. Responsibility and Manipulation. **The Journal of Ethics**, v. 8, n. 2, p. 145-177, 2004. DOI: 10.1023/B:JOET.0000018773.97209.84.

FISCHER, J. M. Ultimacy and Alternative Possibilities. **Philosophical Studies**, v. 144, n. 1, p. 15–20, 2009. DOI: 10.1007/s11098-009-9370-1.

FISCHER, J. M. The Zygote Argument Re-mixed. **Analysis,** v. 71, n. 2, p. 267-272, 2011. DOI: 10.1093/analys/anr008.

FISCHER, J. M.; RAVIZZA, M. **Responsibility and Control:** A Theory Of Moral Responsibility. Cambridge: Cambridge University Press, 1998.

FISCHER, J. M.; TODD, P. The Truth about Freedom: A Reply to Merricks. **The Philosophical Review**, v. 120, n. 1, p. 97-115, 2011. DOI: 10.1215/00318108-2010-025.

FISCHER, J. M.; TODD, P. (ed.). **Freedom, Foreknowledge, and Fatalism.** New York: Oxford University Press, 2015.

FLINT, T. P. Compatibilism and the Argument from Unavoidability. **Journal of Philosophy**, v. 74, n. 8, p. 423-440, 1987. DOI: 10.2307/2027000.

FOLEY, R. Compatibilism and Control over the Past. **Analysis**, v. 39, n. 2, p. 70-74, 1979. DOI: 10.1093/analys/39.2.70.

FRANKFURT, H. G. Alternate Possibilities and Moral Responsibility. **Journal of Philosophy**, v. 66, n. 23, p. 820-39, 1969. Reimpresso em: Frankfurt (1988), Watson (2003, cap. 8), Pereboom (2009a, cap. 15) e Russell e Deery (2013, cap. 7). DOI: 10.2307/2023833.

FRANKFURT, H. G. Freedom of the Will and the Concept of a Person. **Journal of Philosophy**, v.68, n. 1, p. 5-20, 1971. Reimpresso em Frankfurt (1988), Watson (2003, cap. 16), Pereboom (2009a: cap. 16) e Russell e Deery (2013, cap. 13). DOI: 10.2307/2024717.

FRANKFURT, H. G. **The Importance of What we Care About**. Cambridge: Cambridge University Press, 1988.

FRANKFURT, H. G. Reply to John Martin Fischer. *In*: BUSS, S.; OVERTON, L. (ed.). **Contours of Agency**: Essays on Themes from Harry Frankfurt. Cambridge, MA: MIT Press, 2002.

FRANKLIN, C. E. Powers, Necessity, and Determinism. **Thought**, v. 3, n.3, p. 225-229. DOI: 10.1002/tht3.139.

FRANKLIN, C. E. If Anyone Should Be an Agent-Causalist, then Everyone Should Be an Agent-Causalist. **Mind**, v. 125, n. 500, p. 1101-1131, 2016. DOI: 10.1093/mind/fzv177.

GALLOIS, A. Van Inwagen on Free Will and Determinism. **Philosophical Studies**, v. 32, n. 1, p. 99-105, 1977. DOI: 10.1007/BF00373718.

GINET, C. Might We Have No Choice? *In*: LEHRER, K. **Freedom and Determinism**. New York: Random House, 1966, p. 87-104. (Random House Studies in Philosophy).

GINET, C. In Defense of Incompatibilism. **Philosophical Studies**, v. 44, n. 3, p. 391– 400, 1983.

GINET, C. **On Action**. Cambridge: Cambridge University Press, 1990.

GINET, C. 1996. In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing. **Philosophical Perspectives**, v. 10, p. 403-417. Reimpresso em Widerker e McKenna (2003, p. 75-90). DOI: 10.2307/2216254.

GRAHAM, P. A. A Defense of Local Miracle Compatibilism. **Philosophical Studies**, v. 140, n. 1, p. 65-82, 2008. DOI: 10.1007/s11098-008-9226-0.

GRZANKOWSKI, A. Can' and the Consequence Argument. **Ratio**, v. 27, n. 2, p. 173-189, 2013. DOI: 10.1111/rati.12033.

GRIFFITH, M.; LEVY, N.; TIMPE, K. **Routledge Companion to Free Will**. New York: Routledge, 2017.

HAJI, I. **Incompatibilism's Allure:** Principal Arguments for Incompatibilism. Ontario: Broadview Press, 2009.

HALL, N. Causation and the Price of Transitivity. **Journal of Philosophy**, v. 97, n. 4, p. 198-222, 2000. DOI: 10.2307/2678390.

HAWTHORNE, J. Freedom in Context. **Philosophical Studies**, v. 104, n. 1, p. 63-79, 2001. DOI: 10.1023/A:1010398805497.

HELLER, M. The Mad Scientist Meets the Robot Cats: Compatibilism, Kinds, and Counterexamples. **Philosophy and Phenomenological Research**, v. 56, n. 2, p. 33-337, 1996. DOI: 10.2307/2108523.

HELLER, M. The Disconnect Problem and the Influence Strategy. *In*: KELLER, J. A. **Being, Freedom, and Method**: Themes from the Philosophy of Peter van Inwagen. Oxford: Oxford Scholarship Online, 2017, p. 95-117. DOI: 10.1093/acprof:oso/9780198715702.003.0006.

HOBART, R. E. Free Will as Involving Determination and Inconceivable without It. **Mind**, v. 63, n. 169, p. 1-27, 1934. Reimpresso em van Inwagen e Zimmerman (2008, p. 420-432).

HOEFER, C. Freedom from the Inside Out. *In*: CALLENDER, C. (ed.). **Time, Reality, and Experience**. Cambridge: Cambridge University Press, 2002.

HOEFER, C. Causal Determinism. *In*: ZALTA, E. N. (ed.) **Stanford Encyclopedia of Philosophy**. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2016. Disponivel em: https://plato.stanford.edu/archives/spr2016/entries/determinism-causal/. Acesso em: 12 set. 2021.

HOLBACH, B. The Illusion of Free Will. In: HOLBACH, B. **Système de la nature**. Traduzido por H.D. Robinson, 1770. Reimpresso em: KOLAK, D.; MARTIN, R. (ed.). **The Experience of Philosophy**. 5. ed. Oxford: Oxford University Press, 2002, p. 176-181.

HOLLIDAY, W. H. Freedom and the Fixity of the Past. **The Philosophical Review**, v. 121, n. 2, p. 179-207, 2002. DOI: 10.1215/00318108-1539080.

HOLLIDAY, W. H. Freedom and Modality. *In*: KELLER, J. A. **Being, Freedom, and Method**: Themes from the Philosophy of Peter van Inwagen. Oxford: Oxford Scholarship Online, 2017, p. 149-156. DOI: 10.1093/acprof:oso/9780198715702.003.0009.

HOLTON, R. **Willing, Wanting, Waiting**. Oxford: Clarendon Press, 2009. DOI: 10.1093/acprof:oso/9780199214570.001.0001.

HORGAN, T. 'Could', Possible Worlds, and Moral Responsibility. **Southern Journal of Philosophy**, v. 17, n. 3: 345-358, 1979. DOI: 10.1111/j.2041-6962.1979.tb00250.x.

HORGAN, T. Compatibilism and the Consequence Argument. **Philosophical Studies**, v. 47, n. 3, p. 339-356, 1985. DOI: 10.1007/BF00355208.

HORWICH, P. **Asymmetries in Time**. Cambridge, MA: MIT Press, 1987.

HUEMER, M. Van Inwagen's Consequence Argument. **The Philosophical Review**, v. 109, n. 4, p. 525-544, 2000. DOI: 10.1215/00318108-109-4-525.

HUEMER, M. Elusive Freedom? A Reply to Helen Beebee. **The Philosophical Review**, v. 113, n. 3, p. 411-416, 2004. DOI: 10.1215/00318108-113-3-411.

HUME, D. **An Enquiry Concerning Human Understanding**. Editado por Peter Millican. New York: Oxford University Press, 2007.

KANE, R. Two Kinds of Incompatibilism. **Philosophy and Phenomenological Research**, v. 50, n. 2, p. 219-254, 1989. Reimpresso em: O'CONNOR, T. (ed.). **Agents, Causes, and Events**: Essays on Indeterminism and Free Will, New York: Oxford University Press, 1995b, p. 115-150. DOI: 10.2307/2107958.

KANE, R. **The Significance of Free Will**. Oxford: Oxford University Press, 1996. DOI: 10.1093/0195126564.001.0001.

KANE, R. Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism. **Journal of Philosophy**, v. 96, n. 5, p. 217-240, 1999. Reimpresso em: Watson (2003, cap. 15), Russell e Deery (2013, cap. 10). DOI: 10.2307/2564666.

KANE, R. Incompatibilism. *In*: SIDER, T.; HAWTHORNE, J.; ZIMMERMAN, D. W. (ed.). **Contemporary Debates in Metaphysics**. Malden: Blackwell Publishing, 2008.

KANE, R. Rethinking Free Will: New Perspectives on an Ancient Problem, 2011a. *In*: KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b. DOI: 10.1093/oxfordhb/9780195399691.001.0001.

KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b. DOI: 10.1093/oxfordhb/9780195399691.001.0001.

KAPITAN, T. Ability and Cognition: a Defense of Compatibilism. **Philosophical Studies**, v. 63, n. 2, p. 231-43, 1991. DOI: 10.1007/BF00381690.

KAPITAN, T. Incompatibilism and Ambiguity in the Practical Modalities. **Analysis**, v. 56, n. 2, p. 102-110, 1996. DOI: 10.1093/analys/56.2.102.

KAPITAN, T. Autonomy and Manipulated Freedom. *In*: TOMBERLIN, J. (ed.). **Action and Freedom**. Malden: Blackwell Publishing, 2000, p. 81-103.

KAPITAN, T. A Compatibilist Reply to the Consequence Argument. *In*: KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b, p. 131-146. DOI: 10.1093/oxfordhb/9780195399691.001.0001.

KEARNS, S. Aborting the Zygote Argument. **Philosophical Studies**, v. 160, n. 3, p. 379-389, 2012. DOI: 10.1007/s11098-011-9724-3.

KELLER, J. A. (ed.). **Being, Freedom, and Method: Themes from the Philosophy of Peter van Inwagen**. New York: Oxford University Press, 2017. DOI: 0.1093/acprof:oso/9780198715702.001.0001.

KING, M. The Problem with Manipulation. **Ethics**, v. 124, n. 1, p. 65-83, 2013. DOI: 10.1086/671391.

KIM, J. **Mind in a Physical World**: An Essay on the Mind-Body Problem and Mental Causation. Cambridge, MA: Bradford Books, MIT Press, 1998.

KRATZER, A. What 'must' and 'can' must and can mean. **Linguistics and Philosophy**, v. 1, n. 3, p. 337-355, 1997. DOI: 10.1007/BF00353453.

KREMER, M. How Not to Argue for Incompatibilism. **Erkenntnis**, v. 60, n. 1, p. 1-26, 2004. DOI: 10.1023/B:ERKE.0000005087.33155.fc.

LAMB, J. W. On a Proof of Incompatibilism. **The Philosophical Review**, v. 86, n. 1, p. 20–35, 1977. DOI: 10.2307/2184160.

LAMB, J. W. Evaluative Compatibilism and the Principle of Alternative Possibilities. **Journal of Philosophy**, v. 90, n. 1, p. 497-516, 1993. DOI: 10.2307/2941025.

LEHRER, K. (ed.). **Freedom and Determinism**. New York: Random House, 1966.

LEHRER, K. Cans Without Ifs. **Analysis**. v. 29, n. 1, p. 29-32, 1968. DOI: 10.1093/analys/29.1.29.

LEHRER, K. 'Can' in Theory and Practice: A Possible Worlds Analysis. *In*: BRAND, M.; WALTON, D. (ed). **Action theory:** Proceedings of the Winnipeg Conference on Human Action, held at Winnipeg. Dordrecht: Reidel, 1976, p. 241-270

LEHRER, K. Preferences, Conditionals, and Freedom. *In*: VAN INWAGEN, P. (ed.). **Time and Cause:** Essays Presented to Richard Taylor. Dordrecht: Reidel, 1980, p. 187-201. DOI: 10.1007/978-94-017-3528-5_11.

LEHRER, K. Freedom and the Power of Preference. *In*: CAMPBELL, J. K.; O'ROURKE, M.; SHIER, D. (ed.). **Freedom and Determinism**. Press: Bradford Books, 2004, p. 47-69.

LEVY, N. **Hard Luck: How Luck Undermines Freedom**. Oxford: Oxford University Press, 2011. DOI: 10.1093/acprof:oso/9780199601387.001.0001.

LEWIS, D. **Counterfactuals**. Cambridge, MA: Harvard University Press, 1973.

LEWIS, D. The Paradoxes of Time Travel. **American Philosophical Quarterly**, v. 13, n. 2, p. 145-52, 1976.

LEWIS, D. Counterfactual Dependence and Time's Arrow. **Noûs**, v. 13, n. 4, p. 455-476, 1976. DOI: 10.2307/2215339.

LEWIS, D. Are We Free to Break the Laws? **Theoria**, v. 4, n. 3, p. 113-21, 1981. Reimpresso em Watson (2003, cap. 6). DOI: 10.1111/j.1755-2567.1981.tb00473.x.

LEWIS, D. Against Overlap. *In*: LEWIS, D. **On the Plurality of Worlds**. Oxford: Blackwell, 1986, p. 198-209.

LEWIS, D. Finkish Dispositions. **Philosophical Quarterly**, v. 47, n. 187, p. 143-158, 1997. DOI: 10.1111/1467-9213.00052.

LEWIS, D. **On the Plurality of Worlds**. Oxford: Blackwell, 1986.

LEWIS, P. J. **Quantum Ontology: A Guide to the Metaphysics of Quantum Mechanics**. New York: Oxford University Press, 2016. DOI: 10.1093/acprof:oso/9780190469825.001.0001.

LIBET, B. Do We Have Free Will? **Journal of Consciousness Studies**, v. 6, n. 8-9, p. 47-57, 1999. Reimpresso em Sinnott-Armstrong e Nadel (2011, p. 1-10).

LIST, C. Free Will, Determinism, and the Possibility of Doing Otherwise. **Noûs**, v. 48, n. 1, p. 156-178, 2014. DOI: 10.1111/nous.12019.

LOEWER, B. Humean Supervenience. **Philosophical Topics**, v. 24, n. 1, p. 101-27, 1996a. DOI: 10.5840/philtopics199624112.

LOEWER, B. Freedom from Physics: Quantum Mechanics and Free Will. **Philosophical Topics**, v. 24, n. 2, p. 91-112, 1996b. DOI: 10.5840/philtopics19962428.

LYCAN, W. G. Free Will and the Burden of Proof. **Royal Institute of Philosophy Supplements**, v. 53, p. 107-122, 2003. DOI: 10.1017/S1358246100008298.

MACKIE, P. Fatalism, Incompatibilism, and the Power to Do Otherwise. **Noûs**, v. 37, n. 4, p. 672-689, 2003. DOI: 10.1046/j.1468-0068.2003.00455.x.

MACKIE, P. Mumford and Anjum on Incompatibilism, Powers, and Determinism. **Analysis**, v. 74, n. 4, p. 593-603, 2014a. DOI: 10.1093/analys/anu088.

MACKIE, P. Counterfactuals and the Fixity of the Past. **Philosophical Studies**, v. 168, n. 2, p. 397-415, 2014b. DOI: 10.1007/s11098-013-0135-5.

MAIER, J. The Agentive Abilities. **Philosophy and Phenomenological Research**, v. 90, n. 1, p. 113-134, 2013. DOI: 10.1111/phpr.12038.

MARKOSIAN, N. A Compatibilist Version of the Theory of Agent Causation. **Pacific Philosophical Quarterly**, v. 80, n. 3, p. 257-277, 1999. DOI: 10.1111/1468-0114.00083.

MCKAY, T. J.; JOHNSON, D. A Reconsideration of an Argument against Compatibilism. **Philosophical Topics**, v. 24, n. 2, p. 113-122, 1996. DOI:

10.5840/philtopics199624219.

MCKENNA, M. Source Incompatibilism, Ultimacy, and the Transfer of Non-Responsibility. **American Philosophical Quarterly**, v. 38, n. 1, p. 37-52, 2001.

MCKENNA, M. Responsibility and Globally Manipulated Agents. **Philosophical Topics**, John Fischer (ed.), v. 32, n. 1-2, p. 169-192, 2004a. Reimpresso em: Russell e Deery (2013, cap. 18). DOI: 10.5840/philtopics2004321/222.

MCKENNA, M. 2004b. Compatibilism. *In*: ZALTA, E. N. **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2015. Disponível em: https://plato.stanford.edu/archives/win2016/entries/compatibilism/. Acesso em: 12 set. 2021.

MCKENNA, M. A Hard-line Reply to Pereboom's Four-Case Manipulation Argument. **Philosophy and Phenomenological Research**, v. 77, n. 1, p. 142-159, 2008. DOI: 10.1111/j.1933-1592.2008.00179.x.

MCKENNA, M. Whose Argumentative Burden, which Incompatibilist Arguments? Getting the Dialectic Right. **Australasian Journal of Philosophy**, v. 88, n. 3, p. 429-443, 2010. DOI: 10.1080/00048400903233811.

MCKENNA, M. Contemporary Compatibilism: Mesh Theories and Reasons: Responsive Theories, 2011. *In*: KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b, p. 175-196. DOI: 10.1093/oxfordhb/9780195399691.001.0001.

MCKENNA, M. **Conversation and Responsibility**. New York: Oxford University Press, 2012. DOI: 10.1093/acprof:oso/9780199740031.001.0001.

MELE, A. R. **Autonomous Agents: From Self-Control to Autonomy**. New York: Oxford University Press, 1995. DOI: 10.1093/0195150430.001.0001.

MELE, A. R. **Free Will and Luck**. New York: Oxford University Press, 2006. DOI: 10.1093/0195305043.001.0001.

MELE, A. R. Manipulation, Compatibilism, and Moral Responsibility. **The Journal of Ethics**, v. 12, n. 3-4, p. 263-286, 2008. DOI: 10.1007/s10892-008-9035-x.

MELE, A. R. **Effective Intentions:** The Power of the Conscious Will. New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780195384260.001.0001.

MELE, A. R. (ed.). **Surrounding Free Will: Philosophy, Psychology, Neuroscience**. Oxford: Oxford University Press, 2015. DOI: 10.1093/acprof:oso/9780199333950.001.0001.

MELE, A. R.; ROBB, D. Rescuing Frankfurt-Style Cases. **The Philosophical Review**, v. 107, n. 1, p. 97-112, 1998. DOI: 10.2307/2998316.

MERRICKS, T. Truth and Freedom. **The Philosophical Review**, v. 118, n.1, p. 29-257, 2009. DOI: 10.1215/00318108-2008-028.

MICKELSON, K. [DEMETRIOU, K.]. The Soft-Line Solution to Pereboom's Four-Case Argument. **Australasian Journal of Philosophy**, v. 88, n. 4, p. 595-617, 2010. doi:10.1080/00048400903382691.

MICKELSON, K. [DEMETRIOU, K.]. A Critique of Vihvelin's Three-Fold Classification. **Canadian Journal of Philosophy**, v. 45, n. 1, p. 85-99, 2015a. DOI: 10.1080/00455091.2015.1009321.

MICKELSON, K. [DEMETRIOU, K.]. The Zygote Argument is Invalid: Now What? **Philosophical Studies**, v. 172, n. 11, p. 2911-2929, 2015b. DOI: 10.1007/s11098-015-0449-6.

MOORE, G. E. **Ethics**. Oxford: Oxford University Press, 1912.

MOORE, G. E. Free Will. *In*: MOORE, G. E. **Ethics**. Oxford: Oxford University Press, 1912, cap. 6.

MUMFORD, S.; ANJUM, R. L. A New Argument Against Compatibilism. **Analysis**, v. 74, n. 1, p. 20-25, 2014. DOI: 10.1093/analys/ant095.

MURRAY, D.; NAHMIAS, E. Explaining Away Incompatibilist Intuitions. **Philosophy and Phenomenological Research**, v. 88, n. 2, p. 434-467, 2014. DOI: 10.1111/j.1933-1592.2012.00609.x.

NAGEL, T. Moral Luck. In: NAGEL, T. **Mortal Questions.** New York: Cambridge University Press, 1979. Reprinted in Russell & Deery 2013: ch. 1.

NAHMIAS, E. The Phenomenology of Free Will. **Journal of Consciousness Studies**, v. 11, n. 7-8, p. 162-179, 2004. Reimpresso em: RUSSELL, P.; DEERY, O. (ed.). **The Philosophy of Free Will, Essential Readings from the Contemporary Debates**. Oxford: Oxford University Press, 2013, cap. 25.

NAHMIAS, E. Intuitions about Free Will, Determinism, and Bypassing, 2011. *In*: KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b, p. 555–573. DOI: 10.1093/oxfordhb/9780195399691.001.0001.

NAHMIAS, E. *et al*. Is Incompatibilism Intuitive? **Philosophy and Phenomenological Research**, v. 73, n. 1, p. 28-53, 2006. DOI: 10.1111/j.1933-1592.2006.tb00603.x.

NELKIN, D. The Consequence Argument and the Mind Argument. **Analysis**, v. 61, n. 2, p. 107-115, 2001. Reimpresso em: RUSSELL, P.; DEERY, O. (ed.). **The Philosophy of Free Will, Essential Readings from the Contemporary Debates**. Oxford: Oxford University Press, 2013, cap. 6. DOI: 10.1093/analys/61.2.107.

NELKIN, D. K. Deliberative Alternatives. **Philosophical Topics**, v. 32, n. 1-2, p. 215-240, 2004. DOI: 10.5840/philtopics2004321/224.

NELKIN, D. K. **Making Sense of Freedom and Responsibility**. Oxford: Oxford University Press, 2011. DOI: 10.1093/acprof:oso/9780199608560.001.0001.

NELKIN, D. K.; RICKLESS, S. Warfield's New Argument for Incompatibilism. **Analysis**, v. 62, n. 2, p. 104-107, 2002. DOI: 10.1093/analys/62.2.104.

NICHOLS, S. The Indeterminist Intuition: Source and Status. **The Monist**, v. 95, n. 2, p. 290-307, 2012.

OAKLEY, S. Defending Lewis's Local Miracle Compatibilism. **Philosophical Studies**, v. 130, n. 2, p. 337-349. DOI: 10.1007/s11098-004-4677-4.

O'CONNOR, T. On the Transfer of Necessity. **Noûs**, v. 27, n. 2, p. 204-218, 1993. DOI: 10.2307/2215756.

O'CONNOR, T. Agent Causation. In: O'CONNOR, T. (ed.). **Agents, Causes, and Events: Essays on Indeterminism and Free Will**. New York: Oxford University Press, 1995b, p. 173-200. Reimpresso em: WATSON, G. (ed.). **Free Will**. 2. ed. Oxford: Oxford University Press, 2003, cap. 13.

O'CONNOR, T. (ed.). **Agents, Causes, and Events: Essays on Indeterminism and Free Will**. New York: Oxford University Press, 1995b.

O'CONNOR, T. **Persons and Causes:** the Metaphysics of Free Will. New York: Oxford University Press, 2000. DOI: 10.1093/019515374X.001.0001.

O'CONNOR, T. Agent-Causal Theories of Freedom, 2011. *In*: KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b, p. 309-327.

PEREBOOM, D. Determinism al Dente. *Noûs*, 29(1): 21–45, 1995. Reimpresso em: PEREBOOM, D. (ed.). **Free Will**. 2. ed. Indianapolis: Hackett Publishing, 2009a, cap. 22. DOI: 10.2307/2215725.

PEREBOOM, D. **Living Without Free Will**. Cambridge: Cambridge University Press, 2001.

PEREBOOM, D. Source Incompatibilism and Alternative Possibilities. *In*: WIDERKER, D.; MCKENNA, M. (ed.). **Moral responsibility and alternative possibilities**. Burlington: Ashgate Publishing Company, 2003, p. 185-200.

PEREBOOM, D. Defending Hard Incompatibilism. **Midwest Studies in Philosophy**, v. 29, n. 228-247, 2005. DOI: 10.1111/j.1475-4975.2005.00114.x.

PEREBOOM, D. A Hard-line Reply to the Multiple-Case Manipulation Argument. **Philosophy and Phenomenological Research**, v. 77, n. 1, p. 160-170, 2008. DOI: 10.1111/j.1933-1592.2008.00192.x.

PEREBOOM, D. (ed.). **Free Will**. 2. ed. Indianapolis: Hackett Publishing, 2009a.

PEREBOOM, D. Hard Incompatibilism and its Rivals. **Philosophical Studies**, v. 144, n. 1, p. 21-33, 2009b. DOI: 10.1007/s11098-009-9371-0.

PEREBOOM, D. **Free Will, Agency, and Meaning in Life**. Oxford: Oxford University Press, 2014. DOI: 10.1093/acprof:oso/9780199685516.001.0001.

PENDERGRAFT, G. The Explanatory Power of Local Miracle Compatibilism. **Philosophical Studies**, v. 156, n. 2, p. 249-266. DOI: 10.1007/s11098-010-9594-0.

PERRY, J. Compatibilist Options. *In*: CAMPBELL, J. K.; O'ROURKE, M.; SHIER, D. (ed.). **Freedom and Determinism**. Press: Bradford Books, 2004, p. 231-254.

PERRY, J. Can't We All Just be Compatibilists? A Critical Study of John Martin Fischer's My Way. **The Journal of Ethics**, v. 12, n. 2, p. 157-166, 2008. DOI: 10.1007/s10892-008-9030-2.

PETTIT, P.; SMITH, M. Freedom in Belief and Desire. **Journal of Philosophy**, v. 93, n. 9, p. 429-449, 1996. DOI: 10.2307/2940892.

ROSEN, G. The Case for Incompatibilism. **Philosophy and Phenomenological Research**, v. 64, n. 3, p. 699-706, 2002. DOI: 10.1111/j.1933 1592.2002.tb00168.x.

ROSKIES, A.; NAHMIAS, E. 'Local Determination', even if we could find it, does not challenge free will: Commentary on Marcelo Fischborn. **Philosophical Psychology**, v. 30, n. 1-2, p. 185-197. DOI: 10.1080/09515089.2016.1248286.

RUSSELL, P. **Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility**. New York: Oxford University Press, 1995.

RUSSELL, P.; DEERY, O. (ed.). **The Philosophy of Free Will, Essential Readings from the Contemporary Debates**. Oxford: Oxford University Press, 2013.

SARTORIO, C. Causes as Difference-Makers. **Philosophical Studies**, v. 123, n. 1-2, p. 71-96, 2005. DOI: 10.1007/s11098-004-5217-y.

SARTORIO, C. **Causation and Free Will**. Oxford: Oxford University Press, 2016. DOI: 10.1093/acprof:oso/9780198746799.001.0001.

SCANLON, T.M. **Moral Dimensions: Permissibility, Meaning, Blame**. Cambridge, MA: Belknap Press of Harvard University Press, 2008.

SCHAFFER, J. Causation and Laws of Nature: Reductionism. *In*: SIDER, T.; HAWTHORNE, J.; ZIMMERMAN, D. W. (ed.). **Contemporary Debates in Metaphysics**. Malden: Blackwell Publishing, 2008, p. 82-107.

SCHLOSSER, M. E. Manipulation and the Zygote Argument: Another Reply. **The Journal of Ethics**, v. 19, n. 1, p. 73-84, 2015. DOI: 10.1007/s10892-014-9183-0.

SCHLOSSER, M. E. Traditional Compatibilism Reformulated and Defended. **Journal of Philosophical Research**, online 30 jun. 2017. (no prelo). DOI: 10.5840/jpr2017629108.

SCHNEIDER, B. S. Compatibilism and the Notion of Rendering Something False. **Philosophical Studies**, v. 117, n. 3, p. 409-428, 2004. DOI: 10.1023/B:PHIL.0000016484.52149.52.

SCHNEIDER, B. S. On What We Can Ensure. **Synthese**, v. 162, n. 1, p. 101-115, 2008. DOI: 10.1007/s11229-007-9175-8.

SEHON, S. Action Explanation and the Free Will Debate: How Incompatibilist Arguments Go Wrong. **Philosophical Issues**, v. 22, n.1, p. 351-368, 2012. DOI: 10.1111/j.1533-6077.2012.00234.x.

SEHON, S. **Free Will and Action Explanation**. Oxford: Oxford University Press, 2016. DOI: 10.1093/acprof:oso/9780198758495.001.0001.

SHABO, S. Uncompromising Source Incompatibilism", **Philosophy and Phenomenological Research**, v. 80, n.2, p. 349-383, 2010. DOI: 10.1111/j.1933-1592.2010.00328.x.

SHOEMAKER, S. **Identity, Cause, and Mind**: Philosophical Essays. London: Cambridge University Press, 1984.

SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. Vol. 1. Oxford: Oxford University Press, 2013. DOI: 10.1093/acprof:oso/9780199694853.001.0001.

SHOEMAKER, S. Causality and Properties. *In*: SHOEMAKER, S. **Identity, Cause, and Mind**: Philosophical Essays. London: Cambridge University Press, 1984.

SIDER, T. **Four-Dimensionalism:** An Ontology of Persistence and Time. Oxford: Clarendon Press, 2001. DOI: 10.1093/019924443X.001.0001.

SIDER, T.; HAWTHORNE, J.; ZIMMERMAN, D. W. (ed.). **Contemporary Debates in Metaphysics**. Malden: Blackwell Publishing, 2008.

SINNOTT-ARMSTRONG, W.; NADEL, L. **Conscious Will and Responsibility:** A Tribute to Benjamin Libet. New York: Oxford University Press, 2011. DOI: 10.1093/acprof:oso/9780195381641.001.0001.

SLOTE, M. Selective Necessity and the Free Will Problem. **Journal of Philosophy**, v. 79, n. 1, p. 5-24, 2001. DOI: 10.2307/2026343.

SMILANSKY, S. **Free Will and Illusion**. Oxford: Clarendon Press, 2000.

SMITH, M. **Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics**. New York: Cambridge University Press, 2004.

SMITH, M. A Theory of Freedom and Responsibility. *In*: CULLITY, G.; GAUT, B. (ed.). **Ethics and Practical Reason**. New York: Clarendon Press, 1997, p. 293-317.

SMITH, M. Rational Capacities. *In*: SMITH, M. **Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics**. New York: Cambridge University Press, 2004, p. 114-135. DOI: 10.1017/CBO9780511606977.008.

SPEAK, D. The Consequence Argument Revisited. 2011. *In*: KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b, p. 115-128.

SPENCER, J. Able to Do the Impossible. **Mind**, v. 126, n. 502, p. 466-497, 2017. DOI: 10.1093/mind/fzv183.

SPITZLEY, J. Why Pereboom's Four-Case Manipulation Argument is Manipulative. **Journal of Cognition and Neuroethics**, v. 3, n. 1, p. 363-382. Disponível em: http://jcn.cognethic.org/jcnv3i1_Spitzley.pdf. Acesso em: 15 set. 2021.

SRIPADA, C. S. What Makes a Manipulated Agent Unfree? **Philosophy and Phenomenological Research**, v. 85, n. 3, p. 563-593, 2012. DOI: 10.1111/j.1933-1592.2011.00527.x.

STEWARD, H. Fairness, Agency, and the Flicker of Freedom. **Noûs**, v. 43, n. 1, p. 64-93, 2009. DOI: 10.1111/j.1468-0068.2008.01696.x.

STEWARD, H. **A Metaphysics for Freedom**. Oxford: Oxford University Press, 2012. DOI: 10.1093/acprof:oso/9780199552054.001.0001.

STRAWSON, G. **Freedom and Belief**. Oxford: Clarendon Press, 1986. DOI: 10.1093/acprof:oso/9780199247493.001.0001.

STRAWSON, G. The Impossibility of Ultimate Moral Responsibility. **Philosophical Studies**, v. 75, n. 1-2, p. 5-24, 1994. Reimpresso em: Pereboom (2009a, cap. 21) e Russell e Deery (2013, cap. 19). DOI: 10.1007/BF00989879.

STRAWSON, P. F. Freedom and Resentment. **Proceedings of the British Academy**, v. 48, p. 187-211, 1962. Reimpresso em: Watson (2003, cap. 4) e Pereboom (2009a, cap. 13).

SWARTZ, N. **The Concept of Physical Law**. Cambridge: Cambridge University Press, 1986.

TAYLOR, C.; DENNETT, D. Who's Still Afraid of Determinism? Rethinking Causes and Possibilities. 2002. *In*: KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b, p. 221-240.

TAYLOR, R. Fatalism. **The Philosophical Review**, v. 71, n. 1, p. 56-66, 1962. DOI: 10.2307/2183681.

TAYLOR, R. **Metaphysics**. 4. ed. Englewood Cliffs, NJ: Prentice-Hall, 1992.

TODD, P. Defending a (modified version of) the Zygote Argument. **Philosophical Studies**, v. 164, n. 1, p. 189-203. DOI: 10.1007/s11098-011-9848-5.

TOGNAZZINI, N. A. The Structure of a Manipulation Argument. **Ethics**, v. 124, n. 2, p. 358-369, 2014. DOI: 10.1086/673434.

TOGNAZZINI, N. A.; FISCHER, J. M. Incompatibilism and the Fixity of the Past. *In*: KELLER, J. A. (ed.). **Being, Freedom, and Method: Themes from the Philosophy of Peter van Inwagen**. New York: Oxford University Press, 2017, p. 140-148. DOI: 10.1093/acprof:oso/9780198715702.003.0008.

TOMBERLIN, J. (ed.). **Action and Freedom**. Malden: Blackwell Publishing, 2000.

VAN INWAGEN, P. The Incompatibility of Free Will and Determinism. **Philosophical Studies**, v. 27, n. 3, p. 185-199, 1975. DOI: 10.1007/BF01624156.

VAN INWAGEN, P. (ed.). **Time and Cause:** Essays Presented to Richard Taylor. Dordrecht: Reidel, 1980. DOI: 10.1007/978-94-017-3528-5.

VAN INWAGEN, P. **An Essay on Free Will**. Oxford: Clarendon Press, 1983.

VAN INWAGEN, P. The Mystery of Metaphysical Freedom. *In:* VAN INWAGEN, P.; ZIMMERMAN, D. W. (ed.). **Metaphysics:** The Big Questions. Oxford: Blackwell Publishing, 1998, p. 365-373.

VAN INWAGEN, P. Free Will Remains a Mystery. *In:* TOMBERLIN, J. (ed.). **Action and Freedom**. Malden: Blackwell Publishing, 2000, p. 1-19.

VAN INWAGEN, P. Freedom to Break the Laws. **Midwest Studies in Philosophy**, v. 28, p. 334-350, 2004a. DOI: 10.1111/j.1475-4975.2004.00099.x.

VAN INWAGEN, P. Van Inwagen on Free Will. 2004b. *In:* CAMPBELL, J. K.; O'ROURKE, M.; SHIER, D. (ed.). **Freedom and Determinism**. Press: Bradford Books, 2004, p. 213-230.

VAN INWAGEN, P. How to Think about the Problem of Free Will. **The Journal of Ethics**, v. 12, n. 3-4, p. 327-341, 2008. DOI:10.1007/s10892-008-9038-7.

VAN INWAGEN, P. A Promising Argument. *In:* KANE, R. (ed.). **The Oxford Handbook of Free Will**. 2. ed. Oxford: Oxford University Press, 2011b, p. 475-481. DOI:10.1093/oxfordhb/9780195399691.001.0001.

VAN INWAGEN, P.; ZIMMERMAN, D. W. (ed.). **Metaphysics:** The Big Questions. Oxford: Blackwell Publishing, 1998.

VARGAS, M. Libertarianism and Skepticism about Free Will: Some Arguments against Both. **Philosophical Topics**, v. 32, n. 1-2, p. 403-426, 2004. DOI: 10.5840/philtopics2004321/210.

VARGAS, M. 2013, “How to Solve the Problem of Free Will”, in Russell & Deery 2013: ch. 21.

VARGAS, M. **Building Better Beings**: A Theory of Moral Responsibility. Oxford: Oxford University Press, 2013. DOI: 10.1093/acprof:oso/9780199697540.001.0001.

VIHVELIN, K. The Modal Argument for Incompatibilism. **Philosophical Studies**, v. 53, n. 2, p. 227-244, 1988. DOI: 10.1007/BF00354642.

VIHVELIN, K. Stop Me Before I Kill Again. **Philosophical Studies**, v. 75, n. 1-2, p. 115-148, 1994. DOI: 10.1007/BF00989883.

VIHVELIN, K. What Time Travelers Cannot Do. **Philosophical Studies**, v. 81, n. 2-3, p. 315-330, 1996. DOI: 10.1007/BF00372789.

VIHVELIN, K. Free Will Demystified: A Dispositional Account. **Philosophical Topics**, v. 32, n. 1-2, p. 427-450, 2004. DOI: 10.5840/philtopics2004321/211.

VIHVELIN, K. Compatibilism, Incompatibilism, and Impossibilism. *In:* SIDER, T.; HAWTHORNE, J.; ZIMMERMAN, D. W. (ed.). **Contemporary Debates in Metaphysics**. Malden: Blackwell Publishing, 2008.

VIHVELIN, K. How to Think about the Free Will/Determinism Problem. *In:* CAMPBELL, J. K.; O’ROURKE, M.; SLATER, M. H. (ed.). **Carving Nature at its Joints:** Natural Kinds in Metaphysics and Science. Cambridge: MIT Press, 2011, p. 313-340. DOI: 10.7551/mitpress/9780262015936.003.0014.

VIHVELIN, K. **Causes, Laws, and Free Will:** Why Determinism Doesn’t Matter. New York: Oxford University Press, 2013. DOI: 10.1093/acprof:oso/9780199795185.001.0001.

VIHVELIN, K. How Not to Think about Free Will. **Journal of Cognition and Neuroethics**, v. 3, n. 2, p. 393-403, 2015.

VIHVELIN, K. Dispositional Compatibilism. *In:* GRIFFITH, M.; LEVY, N.; TIMPE, K. **Routledge Companion to Free Will**. New York: Routledge, 2017.

WALLACE, R. J. **Responsibility and the Moral Sentiments**. Cambridge: Harvard University Press, 1994.

WALLER, R. R. The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument. **Philosophia**, v. 42, n. 1, p. 209–222, 2014. DOI: 10.1007/s11406-013-9476-8.

WARFIELD, T. A. Causal Determinism and Human Freedom are Incompatible: A New Argument for Incompatibilism. *In:* TOMBERLIN, J. (ed.). **Action and Freedom**. Malden: Blackwell Publishing, 2000, p. 167-180.

WARFIELD, T. A. Compatibilism and incompatibilism: Some Arguments. *In:* LOUX, M. J.; ZIMMERMAN, D. W. (ed.). **The Oxford Handbook of Metaphysics**. Oxford: Oxford University Press, 2003, p. 614-630.

WATSON, G. Free Agency. **Journal of Philosophy**, v. 72, n. 8, p. 205-220, 1975. DOI: 10.2307/2024703.

WATSON, G. Free Action and Free Will. **Mind**, v. 96, n. 382, p. 145-172, 1987. DOI: 10.1093/mind/XCVI.382.145.

WATSON, G. Two Faces of Responsibility. **Philosophical Topics**, v. 24, n. 2, p. 227-248, 1996. DOI: 10.5840/philtopics199624222.

WATSON, G. (ed.). **Free Will**. 2. ed. Oxford: Oxford University Press, 2003.

WATSON, G. **Agency and Answerability**: Selected Essays. Oxford: Clarendon Press, 2004. DOI: 10.1093/acprof:oso/9780199272273.001.0001.

WEGNER, D. **The Illusion of Conscious Will**. Cambridge: Bradford Books MIT Press, 2003.

WESTPHAL, J. A New Way with the Consequence Argument. **Analysis,** v. 63, n. 3, p. 209-212, 2003. DOI: 10.1093/analys/63.3.208.

WESTPHAL, J. The Future and the Truth-Value Links: a Common Sense View. **ANALYSIS**, v. 66, p. 1-9, 2006. DOI: 10.1093/analys/66.1.1.

WESTPHAL, J. Is There a Modal Fallacy in van Inwagen's 'First Formal Argument'? **Analysis**, v. 72, n. 1, p. 36-41, 2012. DOI: 10.1093/analys/anr138.

WHITTLE, A. Ceteris Paribus, I Could Have Done Otherwise. **Philosophy and Phenomenological Research**, v. 92, n. 1, p. 73-85, 2016. DOI: 10.1111/phpr.12111.

WIDERKER, D. On an Argument for Incompatibilism. **Analysis**, v. 47, n. 1, p. 37-41, 1987. DOI:10.1093/analys/47.1.37

WIDERKER, D. Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. **The Philosophical Review**, v. 104, n. 2, p. 247-261, 1995. DOI:10.2307/2185979

WIDERKER, D.; MCKENNA, M. (ed.). **Moral responsibility and alternative possibilities**. Burlington: Ashgate Publishing Company, 2003.

WIGGINS, D. Towards a Reasonable Libertarianism. *In:* HONDERICH, T. (ed.). **Essays on Freedom of Action.** London: Routledge & Kegan Paul, 1973.

WOLF, S. Sanity and the Metaphysics of Responsibility. *In:* SCHOEMAN, F. D. (ed.). **Responsibility, Character, and Emotions**. Cambridge: Cambridge University Press, 1987. p. 46-62.

WOLF, S. **Freedom within Reason**. Oxford: Oxford University Press, 1990.

(II) Responsabilidade Moral*

Autor: Matthew Talbert
Tradutores: Rhamon de Oliveira Nunes & Vinícius Cezar Bianchi
Revisores Científicos: Karenina do Nascimento Rodrigues, Pedro Merlussi & Kherian Gracher

Uma parte fundamental das nossas práticas morais cotidianas, assim como das nossas relações interpessoais, consiste em julgar se uma pessoa é moralmente responsável pelo seu comportamento. Consideramos as outras pessoas, e também nós mesmos, como responsáveis pelas próprias ações e pelas consequências delas.

O julgamento de que uma pessoa é moralmente responsável por seu comportamento envolve, ao menos à primeira vista, atribuir-lhe certos poderes e capacidades, além de entender o seu comportamento como o resultado (apropriado) do fato de que a pessoa tem e exerce tais poderes e capacidades. Qualquer que seja a teoria correta dos poderes e capacidades em questão (e apurar as diferentes teorias é o objetivo desta introdução), o fato de uma agente possuí-los a qualifica

---

* TALBERT,M. Moral Responsibility. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2019. Disponível em: https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/. Acesso em: 01 ago. 2021.

The following is the translation of the entry on Scientific Method, 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/win2019/entries/moral-responsibility/>. 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/moral-responsibility/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry.

como moralmente responsável num sentido geral: isto é, como aquela que pode ser moralmente responsável pelos exercícios particulares de suas ações. Seres humanos, que são adultos normais, podem possuir os poderes e capacidades em questão, enquanto animais não-humanos, crianças pequenas e aqueles que sofrem de graves deficiências durante seu desenvolvimento, ou mesmo demência (para citar alguns exemplos), são geralmente considerados como desprovidos de tais poderes e capacidades.

Considerar alguém como responsável envolve, novamente, à primeira vista, reagir a esta pessoa de acordo com o julgamento de que ela é moralmente responsável. Tais reações, frequentemente, constituem instâncias de elogio ou culpa morais (embora possa haver razões para permitir um comportamento moralmente responsável que não seja nem elogiável, nem condenável.). A esse respeito, consulte McKenna (2012, p. 16-17) e M. Zimmerman (1988. 61-62). A culpa é a reação que pode ser acompanhada do julgamento de que a pessoa é moralmente responsável pela conduta que é ruim ou errado, ao passo que o elogio é a reação que pode ser acompanhada do julgamento no qual a pessoa é moralmente responsável pelo comportamento que é certo ou bom.

Antes de tudo, devemos salientar que o esquema acima, embora útil, pode ser enganador em certos aspectos, porque, por um lado, sugere uma correspondência e uma simetria entre o elogio e a culpa, as quais podem não existir. Tanto o elogio quanto a culpa são certamente assimétricos no sentido em que se dá muito mais atenção à culpa do que ao elogio. Uma razão é que a culpabilidade, ao contrário do ato de elogiar, é frequentemente entendida como sujeita a uma sanção; por isso que pode parecer mais urgente aos teóricos e teóricas do assunto articular as condições de culpabilidade. Pode ser que, pelas razões mencionadas, exista uma linguagem mais rica para expressar a culpa do que o elogio (WATSON, 2004, p. 283), e a palavra "culpa" apareça em expressões idiomáticas para as quais não há um emprego paralelo imediato de "elogio". Compare as sentenças "S é **culpado** por X" e "S é **elogiado** por X". Note também que "considerar responsável" não é em si uma expressão neutra, pois ela tipicamente surge em contextos de culpa (WATSON, 2004, p. 284). Além disso, pode haver assimetrias mesmo em contextos nos quais elogio e culpa são adequados: a culpa privada é um fenômeno mais comum do que o elogio privado (COATES; TOGNAZZINI, 2013a), e, enquanto pequenos erros podem razoavelmente merecer culpa, o comportamento minimamente decente muitas vezes parece insuficiente para o elogio. Para estas e outras diferenças entre elogio e culpa, veja Eshleman (2014). Por fim, a conhecida assunção de que

o ato de elogiar e a culpabilidade são simétricos ao menos em termos das capacidades que ambos exigem foi questionada (NELKIN, 2008, 2011; WOLF, 1980, 1990). Como a maioria dos trabalhos sobre a responsabilidade moral, este verbete tenderá a focar no lado negativo do fenômeno. Para saber mais, veja o verbete sobre culpa.

Algumas observações gerais sobre o conceito de responsabilidade moral são necessárias antes de introduzirmos suas concepções particulares. No discurso cotidiano, ouve-se com frequência referências à "responsabilidade moral" das pessoas, em que o ponto em questão indica que a pessoa possui algum dever ou obrigação, alguma **responsabilidade**, que deve cumprir com base em algum padrão. Neste sentido, dizemos, por exemplo, que um(a) advogado(a) tem uma responsabilidade para com o seu ou a sua cliente (comportar-se de certo modo, de acordo com certos padrões). Contudo, este verbete não está preocupado com teorias que especificam as responsabilidades das pessoas, no sentido de saber quais são seus deveres e obrigações, mas com teorias sobre o problema de saber se uma pessoa mantém uma relação correta com suas próprias ações e com as consequências dessas ações, de modo que ela seja devidamente responsabilizada. Infelizmente este verbete também não inclui a discussão de alguns tópicos importantes relacionados à responsabilidade moral, tais como a responsabilidade por omissões (*vide* CLARKE, 2014; FISCHER; RAVIZZA, 1998; NELKIN; RICKLESS, 2017a) ou a responsabilidade coletiva (*vide* o verbete sobre responsabilidade coletiva e os volumes 30 e 38 da *Midwest Studies in Philosophy)*.

Devemos também distinguir a responsabilidade moral da responsabilidade causal. Causação é um tópico complicado, mas em geral é suficientemente claro que uma pessoa é causalmente responsável por — isto é, ela é a (ou uma) causa proeminente de — algum acontecimento ou resultado. Porém, os poderes e as capacidades exigidos para a responsabilidade moral não são idênticos aos poderes causais de uma agente, de modo que não podemos inferir responsabilidade moral a partir de uma atribuição de responsabilidade causal. Crianças pequenas, por exemplo, podem causar certos resultados por meio de uma ação ao mesmo tempo em que deixam de satisfazer os requisitos gerais da responsabilidade moral, caso em que não seria apropriado julgá-las como moralmente responsáveis ou sustentar que sejam moralmente responsáveis pelos resultados dos quais são causalmente responsáveis. E mesmo agentes morais responsáveis, em geral, podem explicar ou defender seu comportamento de modo a pôr em dúvida a responsabilidade moral pelos resultados dos quais são causalmente responsáveis. Suponha que S cause uma explosão ao ligar um interruptor: o fato de que S não tinha razão em esperar

esta consequência ao ligar o interruptor pode colocar em dúvida sua responsabilidade moral (ou ao menos sua culpabilidade) pela explosão, sem alterar sua contribuição causal em questão. Tendo distinguido os diferentes sentidos da responsabilidade, a palavra "responsabilidade" irá se referir à "responsabilidade moral" (no sentido aqui definido) durante o restante deste verbete, a menos que indiquemos o contrário.

Até recentemente, o volumoso trabalho filosófico sobre a responsabilidade moral se conduziu no contexto dos debates sobre o livre-arbítrio, que diziam respeito majoritariamente às várias maneiras pelas quais o determinismo poderia ameaçar o livre-arbítrio e a responsabilidade moral. Uma assunção amplamente inquestionada foi a de que o livre-arbítrio é necessário para a responsabilidade moral, de modo que as questões centrais da responsabilidade moral tinham de ter a ver com os ingredientes do livre-arbítrio e se sua posse era compatível com o determinismo. Contudo, a literatura sobre a responsabilidade moral, recentemente, direcionou-se a questões que vão além das preocupações sobre o determinismo. A maior parte deste verbete lidará com esses últimos aspectos acerca do debate da responsabilidade moral. Entretanto, será mais profícuo começarmos com questões que interseccionam os interesses sobre o livre-arbítrio e a responsabilidade moral.

1. Liberdade, Responsabilidade e Determinismo

Como a agente responsável se relaciona com as suas ações? Que poder ela exerce sobre as suas próprias ações? Uma resposta (parcial) é que o poder relevante é uma forma de controle, e, em particular, uma forma de controle tal que a agente **poderia ter agido de outra forma** em relação à ação em questão. Isso captura uma noção comum do livre-arbítrio, e uma das questões centrais sobre o livre-arbítrio é a de saber se a sua posse (no sentido da habilidade para agir de outro modo) é compatível com o determinismo causal (ou, por exemplo, com a presciência divina — ver o verbete sobre presciência e livre-arbítrio).

Se o determinismo causal for verdadeiro, então a ocorrência de qualquer evento (incluindo os eventos que envolvem a deliberação humana, como escolhas e ações) é tornada inevitável e causalmente necessitada pelos fatos sobre o passado (e sobre as leis da natureza) anteriores à sua ocorrência. Sob tais condições, os fatos sobre o presente e sobre o futuro são fixados unicamente pelos fatos sobre o passado (e sobre as leis da natureza): dados esses fatos precedentes, o presente e o futuro podem se desdobrar apenas de um modo. Veja o verbete sobre determinismo

causal para mais informações.

Se a posse do livre-arbítrio requer uma habilidade para agir de outro modo, então podemos ver por que o livre-arbítrio foi frequentemente considerado como incompatível com o determinismo causal. Um modo de se chegar a essa preocupação **incompatibilista** consiste em focar na maneira pela qual a execução de uma ação deve **depender de uma agente** se a agente possui o tipo de livre-arbítrio exigido para a responsabilidade moral. Como estabelece o influente Argumento da Consequência (GINET, 1966; VAN INWAGEN, 1983, p. 55-105; WIGGINS, 1973), a verdade do determinismo parece acarretar que as ações da agente *não* dependem dela, visto que são consequências inevitáveis de coisas sobre as quais a agente em questão não tem controle. Vejamos um resumo informal deste argumento extraído do importante livro de Peter van Inwagen, *An essay on Free Will* (1983, p. 16):

> Se o determinismo for verdadeiro, então os nossos atos são as consequências das leis da natureza e dos eventos em um passado remoto. Mas não depende de nós o que se passou antes de nascermos, nem depende de nós o que as leis da natureza são. Portanto, as consequências dessas coisas (incluindo nossos atos presentes) também não dependem de nós.

Para saber mais sobre um importante argumento que sugere que o Argumento da Consequência mistura diferentes sentidos em que podemos dizer que as leis da natureza não dependem de nós, veja David Lewis (1981). Para saber mais sobre o incompatibilismo e os argumentos incompatibilistas, veja os verbetes sobre Livre-Arbítrio, Argumentos Incompatibilistas e Teorias Incompatibilistas (não determinísticas) do Livre-Arbítrio, assim como Randolph Clarke (2003).

Os compatibilistas mantêm a ideia de que o livre-arbítrio (e/ou responsabilidade moral) seja possível mesmo em um universo determinístico. Versões do compatibilismo foram defendidas desde os tempos antigos. Por exemplo, os Estoicos, — Crisipo, em particular — argumentaram que a verdade do determinismo não acarreta que as ações humanas sejam inteiramente explicadas por fatores externos às agentes; assim, as ações humanas não são necessariamente explicadas de um modo que é incompatível com a culpa e o elogio. Para visões estoicas sobre a liberdade e o determinismo, confira Bobzien (1998) e Salles (2005). Analogamente, filósofos — e filósofas — no período moderno (como Hobbes e Hume) distinguiram o modo

geral no qual nossas ações são necessitadas caso o determinismo seja verdadeiro das instâncias específicas de necessidade que às vezes são impostas sobre nós por limitações cotidianas em nosso comportamento (e.g., impedimentos físicos que tornam impossível agir como escolhemos). A diferença é que a necessidade envolvida no determinismo é compatível com agentes agindo como escolheram agir: mesmo se o comportamento de S for causalmente determinado, ele ainda pode ser um comportamento que a agente escolheu executar. E, talvez, a habilidade que importa para o livre-arbítrio (e para a responsabilidade) seja justamente a habilidade de agir conforme uma escolha, a qual parece exigir apenas a ausência de limitação externa (e não a ausência do determinismo).

A tradição compatibilista foi transposta para o século XX pelos positivistas lógicos, tais como A. J. Ayer (1954) e Moritz Schlick, em 1930 (1966). Vejamos o modo como Schilick expressou o *insight* compatibilista central em 1930 (baseando-se, particularmente, em Hume):

> Liberdade significa o oposto da compulsão; um homem é **livre** se ele não age sob **compulsão**, e ele é compelido ou não livre quando é impedido externamente... quando está trancado, acorrentado, ou quando alguém o força sob a mira de uma arma a fazer o que, de outra forma, ele não faria (1966, p. 59).

Visto que pressões causais determinísticas nem sempre forçam alguém a "fazer o que de outra forma não faria", a liberdade — ao menos do tipo especificado por Schilick — é compatível com o determinismo.

Uma estratégia compatibilista intimamente relacionada com a anterior, influente no início e meados do século XX, consistiu em oferecer **uma análise condicional da habilidade de agir de outro modo** (AYER, 1954; HOBART, 1934; MOORE 1912). Já para expressões anteriores desta estratégia, verifique Hobbes (1654) e Hume (1748). Como já dissemos, mesmo se o determinismo for verdadeiro, as agentes podem frequentemente agir da meneira que escolhem. E a ideia de que uma agente que executou uma ação A (com base em sua escolha de fazê-lo) poderia ter executado uma ação diferente **na condição de que** (contrariamente ao que de fato ocorreu) ela tivesse escolhido executar outra ação, é igualmente compatível com o determinismo. Mesmo se o comportamento atual de uma pessoa for causalmente determinado por seu passado, ainda poderia ser o caso que, se seu passado tivesse

sido adequadamente diferente (ou seja, se os desejos, intenções, escolhas, etc., da pessoa, tivessem sido diferentes), então, ela teria agido de outro modo. E, talvez, isso seja tudo o que a habilidade de agir de outro modo venha a ser: alguém pode agir de outro modo se for verdadeiro que se tivesse escolhido agir de outro modo, então assim o teria agido.

Entretanto, essa imagem compatibilista está aberta a sérias objeções. Em primeiro lugar, pode-se conceder que a habilidade de que se aja como achar melhor é valiosa e, talvez, relacionada ao tipo de liberdade em questão no debate sobre o livre-arbítrio, mas não se segue que isso seja **tudo** o que a posse do livre-arbítrio venha a ser. Uma pessoa que tenha certos desejos como resultado de doutrinação, lavagem cerebral ou psicopatologias pode agir como escolhe, mas sua responsabilidade moral e seu livre-arbítrio ainda podem ser questionados (*vide* §3.2 e §3.3.3). Mais especificamente, a análise condicional está aberta ao seguinte tipo de contraexemplo. Pode ser verdadeiro que uma agente que executa a ação A poderia ter evitado a ação A se assim tivesse escolhido, mas poderia ser verdadeiro também que a agente em questão sofresse de uma compulsão avassaladora por executar A. A análise condicional sugere que a agente conserva a habilidade de agir de outro modo que não seja a ação A, mas, devido à sua compulsão, é claro que lhe falta tal habilidade (BROAD, 1934; CHISHOLM, 1964; LEHRER, 1968, VAN INWAGEN, 1983). De modo mais geral, os(as) incompatibilistas provavelmente ficariam insatisfeitos com a análise condicional, uma vez que ela é incapaz de dar conta de uma habilidade que os agentes possam ter, aqui e agora, para executar ou evitar uma ação ao mesmo tempo em que sustenta que tudo sobre o aqui e o agora, assim como sobre o passado, é fixo.

Apesar das objeções acima, o projeto compatibilista descrito até agora teve uma influência duradoura e significativa. Como veremos abaixo, o fato de que agentes determinadas podem agir como bem entender ainda é uma importante inspiração para os(as) compatibilistas, assim como também é o fato de que agentes determinadas poderiam ter agido diferentemente em situações contrafactuais. Para saber mais, ver o verbete sobre o compatibilismo. Para abordagens mais recentes relacionadas e aprimoradas a partir das abordagens compatibilistas iniciais, confira Michael Fara (2008), Michael Smith (2003), e Kadri Vihvelin (2004), e para uma crítica destas abordagens, ver Randolph Clarke (2009).

Outra tendência influente nos posicionamentos compatibilistas foi a de argumentar que a responsabilidade moral **não** exige uma habilidade de agir de outro modo. Se isto estiver certo, então, o determinismo não ameaçaria a responsabilidade por excluir o

acesso às alternativas comportamentais, embora o determinismo possa ameaçar a responsabilidade de outros modos (VAN INWAGEN 1983, p. 182-88; FISCHER e RAVIZZA, 1998, p. 151-168). Em um artigo muito influente de 1969, Harry Frankfurt oferece exemplos em que pretende mostrar que uma agente pode ser moralmente responsável por uma ação mesmo se ela não pudesse ter agido de outro modo. Versões desses exemplos são muitas vezes chamadas de *Casos de Frankfurt* ou *Exemplos de Frankfurt*. Na forma básica do exemplo, um agente, Jones, delibera sobre se executa uma ação. Outro agente, Black, gostaria que Jones executasse essa mesma ação, e, se necessário, Black poderia **fazer** com que Jones a executasse através de algum tipo de intervenção no processo de deliberação de Jones. Entretanto, no desenrolar dos eventos, Black não intervém na tomada de decisão de Jones, visto que ele pôde perceber que Jones executaria a ação por si próprio e por suas próprias razões. Black não intervém para garantir que a ação de Jones ocorra, mas ele poderia e teria intervido se Jones tivesse dado algum sinal de que não executaria a ação por conta própria. Portanto, **Jones não poderia ter agido de outro modo**, ainda que pareça responsável por seu comportamento. No final das contas, por causa da não intervenção de Black, a ação de Jones é um perfeito caso comum de um comportamento voluntário.

Existem questionamentos sobre se os exemplos de Frankfurt realmente mostram que Jones é moralmente responsável, mesmo que o agente não pudesse ter agido de outro modo. Por um lado, pode não ser claro que Jones realmente não pudesse ter agido de outro modo: apesar de ele ter executado a ação por si próprio, havia a alternativa de tê-la executado devido a alguma intervenção por parte de Black, e não por si próprio. Ademais, ele **poderia** ter dado algum indício a Black que ele não executaria a ação em questão, embora ele não tenha feito isso. Alternativamente, uma objeção poderia ser colocada ao perguntar como Black poderia estar certo de que Jones iria ou não iria executar a ação por si próprio. Parece haver um dilema aqui. Talvez o determinismo ocorra no universo do exemplo e Black perceba alguns sinais que indiquem a presença de fatores que **garantam causalmente** que Jones irá se comportar de determinada maneira. Mas, neste caso, é improvável que os(as) incompatibilistas concedam que Jones seja moralmente responsável, se acaso acreditarem que o determinismo é incompatível com a responsabilidade moral. Por outro lado, talvez o determinismo não seja verdadeiro no universo do exemplo, mas então não fica claro que o exemplo exclua as alternativas para Jones: se o comportamento de Jones não é causalmente determinado, então talvez ele *possa* agir de outro modo. Para objeções deste tipo, aos exemplos originais

de Frankfurt, veja Carl Ginet (1996) e David Widerker (1995); para defesas de Frankfurt, ver John M. Fischer (1994, p. 131-159; 2002, 2010); e para versões refinadas dos exemplos de Frankfurt, negando claramente a Jones o acesso às alternativas, ver Alfred Mele e David Robb (1998), David Hunt (2000) e Derk Pereboom (2000; 2001, p. 18-28).

Em resposta a críticas, como as citadas acima, Frankfurt disse que seu exemplo pretendia sobretudo chamar a atenção ao fato de que "tornar uma ação inevitável **não é a mesma coisa** que fazer com que a ação seja executada." (2006, p. 340. grifo do autor). Particularmente, embora o determinismo possa tornar a ação de uma agente inevitável, disso não se segue que a agente aja **somente** tal qual o faz porque o determinismo é verdadeiro: também pode ser verdadeiro que ela age porque quer e porque vê razões em favor dessa ação. Frankfurt sugere que o ponto de seu exemplo original foi chamar a atenção para a importância das causas **reais** do comportamento que uma agente possa ter (tal como seus desejos e razões) independentemente do fato de saber se a agente pudesse ter feito outra coisa. Ainda conclui que "quando uma pessoa age por suas próprias razões […] a questão de saber se ela poderia ter agido de outro modo é bastante irrelevante" para o propósito de avaliar sua responsabilidade (2006, p. 340). Um foco nas causas reais que levam ao comportamento, tanto quanto a investigação sobre quando se pode dizer que uma agente age sobre **suas próprias** razões, caracterizou muito dos trabalhos sobre a responsabilidade desde o ensaio de Frankfurt (*vide* §2.3 e §3.3.3).

2. Algumas Teorias sobre a Responsabilidade Moral

Esta seção discute três teorias importantes sobre a responsabilidade. As perspectivas adicionais (atribuicionismo, teorias conversacionais, posicionamentos de estrutura ou malhas, posicionamentos céticos etc.) são apresentadas abaixo de maneira relativamente detalhada nas discussões dos debates contemporâneos.

2.1. Teorias Prospectivas

As teorias prospectivas da responsabilidade moral justificam as práticas de responsabilização focando nas consequências benéficas que se pode obter através de tais práticas. Esta teoria foi influente na primeira metade do século XX

(bem como anteriormente), caiu em desuso nas últimas décadas daquele século e recebeu novo interesse mais recentemente.

Tais teorias tendem a enfatizar um dos pontos centrais discutido na seção anterior: uma agente estar sujeita ao determinismo não acarreta que ela esteja sujeita a imposições que a forcem a agir de forma contrária às suas escolhas. Se isto for verdadeiro, então, independentemente da verdade do determinismo, poderia ser útil oferecer incentivos às agentes — elogiá-las e culpá-las, e tratá-las geralmente como responsáveis — a fim de encorajá-las a fazer certas escolhas e, assim, assegurar resultados positivos do seu comportamento.

De acordo com algumas articulações das teorias prospectivas, ser uma agente responsável é simplesmente ser uma agente cujos motivos, escolhas e comportamentos possam ser assim formados. Assim, Moritz Schlick (1966, p. 61, grifo do autor) argumentou que

> A questão de quem é responsável é a questão do **ponto da correta aplicação** do motivo... nisto seu significado está completamente esgotado; por trás dele não se esconde uma conexão misteriosa entre transgressão e represália. É apenas uma questão de saber quem deve ser punido ou recompensado, para que o castigo e a recompensa funcionem como tais - sendo capazes de atingir seu objetivo.

E, de acordo com Schlick, os objetivos da punição e da recompensa não têm a ver com o passado: a ideia de que a punição é "uma **retaliação natural** por um erro passado não deve mais ser defendida numa sociedade instruída" (1966, p. 60, grifo do autor). Ao contrário, a punição deve estar:

> [...] preocupada apenas com a instituição das causas, dos **motivos** de conduta…. Analogamente, no caso da recompensa, estamos preocupados com um incentivo (1966, p. 60, grifo do autor).

J. J. C Smart (1961) também defendeu uma teoria prospectiva da responsabilidade moral bastante conhecida, em meados do século XX. Smart afirmou que culpar uma pessoa pelo seu comportamento é simplesmente avaliá-la negativamente ("desaprová-la", na terminologia de Smart) enquanto simultaneamente

se atribui responsabilidade ao comportamento da agente. E, para Smart, uma atribuição de responsabilidade envolve meramente considerar uma agente como sendo tal que ela teria evitado um comportamento se ela tivesse tido um motivo para executá-lo. Quaisquer que sejam as sanções que podem se seguir em uma atribuição de responsabilidade, elas são administradas com o objetivo de fornecer às agentes motivos para se abster de tal comportamento no futuro.

A teoria geral de Smart tem defensoras e defensores contemporâneos (ARNESON, 2003), mas muitas e muitos encontraram lacunas em aspectos importantes. Uma destas lacunas, como J. Jay Wallace nota, é que "uma teoria como a de Smart deixa de fora o aspecto **atitudinal** subjacente à culpa moral" (WALLACE, 1996, p. 56, grifo do autor). Veja a próxima subseção para saber mais sobre atitudes de culpa. De acordo com Wallace, as atitudes envolvidas na culpa são "retrospectivas e focadas no agente individual que tenha feito algo moralmente errado" (WALLACE, 1996, p. 56). Mas uma teoria prospectiva com foco em provocar os resultados desejados

> [...] não é direcionada exclusivamente para a agente individual que tenha feito algo moralmente errado, mas leva em consideração *qualquer pessoa* suscetível de ser influenciada por nossas reações (WALLACE, 1996, p. 56, grifo do autor).

Em casos excepcionais, um foco nos resultados benéficos pode fornecer fundamentos para tratar como culpáveis as pessoas que sabemos serem inocentes (SMART, 1973). Esta última característica (de algumas) das teorias prospectivas recebeu críticas particularmente fortes.

Esforços recentes foram feitos para desenvolver teorias prospectivas parciais acerca da responsabilidade que evitam algumas das críticas mencionadas acima. Tais teorias (um tanto quanto revisionárias) justificam nossas práticas de responsabilização apelando às suas aptidões no intuito de fomentar a agência moral e a aquisição das capacidades requeridas para tal agência. O "modelo de cultivação de agência responsável", de Manuel Vargas (2013), é bastante notável no que diz respeito a isso. Confira, também, Jefferson (2019) e MacGeer (2015). As recentes abordagens conversacionais da responsabilidade (§3.2.2.) também têm um importante componente prospectivo na medida em que consideram as pessoas, com as quais poderíamos ter interações morais frutíferas, como candidatas à responsabilização. Sobre a responsabilidade moral, alguns céticos também deram ênfase aos benefícios

de certas práticas de responsabilização. Por exemplo, Derk Pereboom — que rejeita a ideia de culpa baseada em merecimento — argumentou que algumas práticas convencionais de culpabilidade podem ser preservadas (mesmo depois que as noções comuns de culpabilidade tenham sido deixadas para trás) na medida em que essas práticas estão fundamentadas em "*desiderata* morais que não invocam merecimento" tais como a "proteção a potenciais vítimas, reconciliação com relacionamentos, tanto pessoais quanto para com a comunidade moral em geral e a formação moral" (2014, p. 134). Confira, também, Caruso (2016), Levy (2012) e Milam (2016). Em contraste a algumas das novas teorias prospectivas descritas acima, Pereboom (2017) propõe que somente aquelas agentes que tenham, de fato, agido imoralmente deveriam estar abertas às práticas de culpabilidade prospectivamente-objetivadas. Para mais informações sobre o ceticismo acerca da responsabilidade, consulte §3.3 e o verbete sobre ceticismo acerca da responsabilidade moral.

2.2. A Teoria das Atitudes Reativas

2.2.1. "Liberdade e Ressentimento"

O artigo de P.F Strawson, de 1962, *Liberdade e Ressentimento*, é um ponto de referência para muitos dos trabalhos sobre a responsabilidade moral que se seguiram a partir dele, especialmente para os(as) compatibilistas. O objetivo de Strawson foi o de traçar uma rota entre as teorias incompatibilistas comprometidas com o requerimento do livre-arbítrio para a responsabilização e as teorias compatibilistas da teoria prospectiva que, de acordo com a visão de Strawson, não reconhecem nem dão conta apropriadamente da importância interpessoal do componente afetivo de nossas práticas de responsabilização moral. Em contraste com as teorias prospectivas como as de J.J.C Smart, Mortiz Schlick (), Strawson foca diretamente nas emoções, as atitudes reativas, que desempenham um papel fundamental em nossas práticas de responsabilização mútua. A sugestão de Strawson é que se dermos atenção à lógica destas respostas emocionais, produziremos uma teoria sobre o que é estar aberto ao elogio e à culpa que não precisa invocar a concepção incompatibilista de livre-arbítrio. Com efeito, a visão de Strawson foi interpretada como sugerindo que nenhum fato metafísico além de nossas práticas de elogio ou culpa é necessário para fundamentar tais práticas.

Parte da inovação da abordagem de Strawson, em 1962, é sua ênfase na "[...] importância que atribuímos às atitudes e intenções de outros seres humanos que são direcionadas a nós". (1993, p. 48). E continua o autor, "[...] no quanto importa para nós se as ações de outras pessoas...refletem atitudes de boa vontade, afeição e estima, por um lado, ou desprezo, indiferença ou malevolência em relação a nós, por outro." (1993, p. 49). Para Strawson, nossas práticas de considerarmos os outros como responsáveis são, em grande medida, reações a estas coisas: isto é, "à qualidade da vontade dos outros em relação a nós." (1993, p. 56).

Para se ter uma ideia da importância da **qualidade da vontade** para nossas relações interpessoais, note a diferença na sua reação para quem lhe fere acidentalmente em comparação a como você reage com alguém que comete a mesma lesão com "desprezo desdenhoso" ou "um desejo malévolo de ferir você" (STRAWSON, 1993, p. 49). O segundo caso provavelmente desperta uma intensidade de ressentimento que não seria, apropriadamente, sentida no primeiro caso. Pontos correspondentes podem ser feitos sobre reações positivas, como a gratidão: você provavelmente não teria o mesmo sentimento de gratidão em relação a uma pessoa que o beneficiou acidentalmente tal como teria com quem o beneficiou por preocupação com o seu bem-estar. O foco aqui está nas atitudes reativas **pessoais**, direcionadas a outrem e que partem de nós, mas Strawson também discute atitudes "derivativas" ou "compreensivas", sentidas para com outras pessoas e "atitudes autorreativas" que uma agente pode direcionar a si mesma.

Na visão de Strawson, a tendência de responder com atitudes reativas relevantes a exibições de boa ou má vontade implica uma demanda por respeito moral e de devida consideração. De fato, para o autor, "a propensão a essas atitudes **é** o que constitui tal demanda", e as atitudes em si mesmas são "correlatas da demanda moral no caso em que se sente que ela é desconsiderada." (1993, p. 63, grifo do autor). Assim, entre as circunstâncias que apaziguam uma atitude de reação (negativa) de alguém, estão aquelas que mostram — a despeito das aparências iniciais — que a demanda pela devida consideração não foi ignorada ou desprezada. Quando uma pessoa explica que a injúria causada a você foi totalmente imprevista e acidental, ela indica que sua consideração pelo seu bem-estar não foi insuficiente e que esta pessoa não é, portanto, um alvo apropriado para atitudes negativas envolvidas na culpa moral.

Note que a agente que isenta a si mesma de culpa do modo exposto acima não está pondo em questão seu estatuto como uma agente responsável: ela ainda está aberta à demanda pela devida consideração e passível, a princípio, a respostas

reativas. Outras agentes, porém, podem não ser alvos legítimos para a culpa e para tais emoções reativas precisamente porque elas não são alvos de uma demanda por consideração. Nesses casos, a agente não está **desculpada**, ela está **isenta** da culpa: não é que se descobriu que seu comportamento não foi malicioso, mas, ao invés disso, ela é vista como alguém de quem não podemos razoavelmente exigir um comportamento melhor. A ampla terminologia usada, na qual o contraste acima é esboçado, "desculpas" *versus* "isenções", deve-se a Watson, de 1987. (2004).

Para Strawson, o grupo mais importante de agentes isentas inclui aquelas que são, ao menos por um tempo, significativamente debilitadas nas relações interpessoais normais. Essas agentes podem ser crianças, ou debilitadas psicologicamente como "esquizofrênicas"; podem exibir "um comportamento puramente compulsivo", ou suas mentes podem ter sido "sistematicamente deturpadas" (STRAWSON 1993, p. 51). Por outro lado, agentes isentas podem simplesmente ser "totalmente carentes…em um sentido moral" (STRAWSON, 1993, p. 58), talvez porque sofreram de "circunstâncias formativas...peculiarmente infelizes" (STRAWSON, 1993, p. 52). Essas agentes não são candidatas à gama de reações emocionais envolvidas em nossas relações pessoais porque não participam nestas relações do modo correto para que aquelas reações lhes sejam aplicadas de maneira sensata. Ao invés de adotar atitudes engajadas-interpessoalmente, que pressupõem uma demanda por respeito em relação a agentes isentas, tomamos uma atitude objetiva em relação a elas. A agente isenta não é considerada "como uma agente moralmente responsável...como um membro da comunidade moral" (STRAWSON, 1993, p. 59), embora ela possa ser considerada como "um objeto de políticas sociais" e como algo a ser "dirigido, ou lidado, ou curado, ou treinado." (STRAWSON, 1993, p. 52).

A perspectiva de Strawson tem uma conclusão compatibilista importante. Podemos ser capazes de, em circunstâncias limitadas, tomar uma perspectiva imparcial, objetiva, sobre o comportamento normal (isto é, as não isentas) de agentes. Mas Strawson argumenta que não podemos tomar essa perspectiva permanentemente e certamente não com base na descoberta de que o determinismo é verdadeiro:

> O compromisso humano com a participação em relacionamentos interpessoais comuns é, penso eu, completa e profundamente enraizado para que levemos a sério o pensamento de que uma convicção

> teórica geral [e.g., sobre a verdade do determinismo] possa mudar nosso mundo e que nele não haveria mais relações interpessoais tais como normalmente as entendemos; e estar envolvido em relacionamentos interpessoais... é precisamente ser exposto à gama de atitudes e sentimentos reativos que estão em questão. (1993, p. 54).

Mais especificamente, a verdade do determinismo não mostraria que os seres humanos geralmente se ocupam de condições desculpáveis ou condições de isenções que tornariam inapropriadas as atitudes envolvidas em tomar uns aos outros como responsáveis. Não se seguiria da verdade do determinismo, por exemplo, "que qualquer ser humano que causasse uma injúria **ou** simplesmente ignora tê-la causado **ou** tem razões superiores aceitáveis para fazê-lo." (STRAWSON, 1993, p. 53, grifo do autor), nem se seguiria (da verdade do determinismo)

> que ninguém saberia o que está fazendo ou que o comportamento das pessoas seja ininteligível em termos de propósitos conscientes, ou que as pessoas vivem em um mundo de desilusão, ou que ninguém tenha um senso moral (STRAWSON, 1993, p. 59).

2.2.2. As Críticas à Teoria de Strawson

Diversas objeções foram levantadas em relação à abordagem teórica geral de P. F. Strawson sobre a responsabilidade moral, suas assunções sobre a psicologia e a sociabilidade humanas e seus argumentos a favor da compatibilidade entre o determinismo e a responsabilidade.

Como notamos na subseção anterior, Strawson argumenta que o conhecimento da verdade do determinismo não levantaria preocupações gerais sobre as nossas práticas de responsabilização, porque a verdade do determinismo não mostraria que seres humanos são **geralmente anormais** de um modo que colocaria em questão sua abertura a atitudes reativas: nenhuma tese que não seja em si mesma contraditória pode ter como consequência que a anormalidade é a condição universal (STRAWSON, 1993, p. 54). Em resposta, observou-se que, apesar da verdade do determinismo poder não sugerir a anormalidade universal, ela pode muito bem mostrar

que seres humanos normais são moralmente incapacitados de um modo que seja relevante para as nossas práticas de responsabilização (RUSSELL, 1992, p. 298-301).

As assunções de Strawson de que somos tão profunda e naturalmente comprometidos e comprometidas com as nossas práticas envolvendo atitudes reativas para desistir delas e que fazê-lo distorceria nossas vidas morais de maneira irreparável também foram criticadas (NELKIN, 2011, p. 42-45; STRAWSON, 1986, p. 84-120; WATSON, 2004, p. 255-258).

Um tipo diferente de objeção enfatiza a **dependência-da-reação** da teoria de Strawson: isto é, o modo pelo qual a teoria explica a responsabilidade da agente em termos de reações morais que caracterizam uma dada comunidade de práticas morais de responsabilização, em vez de fazê-la em termos de fatos independentes sobre se a agente *é* responsável. Essa característica da abordagem de Strawson nos convida a uma leitura que pode ser paradoxal:

> Na visão de Strawson, não há uma noção independente de responsabilidade que explique a adequação das atitudes reativas. A prioridade explicativa se dá ao contrário: não é que tomamos as pessoas como responsáveis porque elas *são* responsáveis; ao contrário, a ideia (**nossa** ideia) de que somos responsáveis deve ser entendida pela prática que, por si só, não é uma questão de sustentar que algumas proposições sejam verdadeiras, mas de expressar nossas preocupações e demandas sobre o tratamento que damos entre nós (WATSON, 2004, p. 222, grifo do autor).[2]

A teoria de Strawson seria particularmente problemática se, como a leitura acima sugere, acarretasse que as práticas de responsabilização de um grupo - tal como se dão e **seja lá como** se dão - estão além da crítica simplesmente porque essas são práticas deste grupo (FISCHER; RAVIZZA, 1993a, p. 18).

Mas há algo a ser dito do outro lado do debate. Parece óbvio que as pessoas sejam apropriadamente consideradas como responsáveis somente se

---

[2] N.T.: Confira Bennett (1980) para uma interpretação não cognitivista relacionada à teoria de Strawson.

existirem fatos independentes sobre sua responsabilidade. Porém, pensando bem, e segundo a influente teoria Strawsoniana de R. Jay Wallace (1996), pode ser difícil "fazer sentido da ideia de um domínio prévio e profundamente independente de fatos sobre a responsabilização moral" que é separado de nossas práticas e, ainda assim, para o qual as nossas práticas devem responder (1996, p. 88). Para Wallace, desistir dos fatos da responsabilidade que são independentes da prática não significa desistir dos fatos sobre a responsabilização; pelo contrário, "devemos interpretar os fatos relevantes sobre a responsabilização como de algum modo dependente de nossas práticas de considerar as pessoas como responsáveis" (1996, p. 89). Tal interpretação requer uma investigação sobre nossas práticas, e, para Wallace, o que emerge mais notoriamente a partir dessa investigação é o grau em que as nossas práticas de responsabilização estão organizadas em torno de um comprometimento fundamental com a justiça (1996, p. 101). Wallace desenvolve este comprometimento com a justiça, e com as normas de justiça, em uma teoria das condições sob as quais as pessoas são apropriadamente consideradas como responsáveis moralmente por seu comportamento (1996, p. 103-109). Para uma defesa mais recente da teoria da **dependência-da-reação** em relação à responsabilidade, consulte Shoemaker (2017b), e, para a crítica de tais teorias, Todd (2016).

2.3. As Teorias da Sensibilidade às Razões

Como visto em §1, uma das influências duradouras da defesa do compatibilismo de Harry Frankfurt foi chamar a atenção para as causas reais do comportamento da agente e, particularmente, para a questão de saber se a agente, mesmo se for causalmente determinada, agiu por suas próprias razões. As teorias da sensibilidade às razões para a responsabilização foram particularmente atenciosas a tais questões.

Essas teorias fundamentam a responsabilização por referência às capacidades da agente para ser apropriadamente sensível a considerações racionais que dão suporte às suas ações. Interpretada de maneira ampla, a teoria da sensibilidade às razões inclui uma coleção diversa de posicionamentos, como o de David Brink e Dana Nelkin (2013), John M. Fischer e Mark Ravizza (1998), Ishtiyaque Haji (1998), Michael McKenna (2013), Dana Nelkin (2011), Carolina Sartorio (2016), R. Jay Wallace (1996) e Susan Wolf (1990). *Responsibility and Control* (1998), de Fischer e Ravizza, que se baseia em Fischer (1994), oferece a articulação mais influente da teoria da sensibilidade às razões.

Fischer e Ravizza começam com uma distinção entre controle regulador e controle de direcionamento. O controle regulador envolve a posse de um poder dual: "o poder de livremente executar algum ato A, e o poder de livremente fazer outra coisa" (1998, p. 31). O controle de direcionamento, por outro lado, não exige acesso a alternativas: ele é manifestado quando uma agente guia seu comportamento em uma direção particular (e independentemente de saber se estava lhe disponível guiar seu comportamento em uma direção diferente). Visto que Fischer e Ravizza tomam o caso de Frankfurt (§ 1) para mostrar que o acesso às alternativas comportamentais não é necessário para a responsabilidade moral, os autores concluem que "o tipo de controle necessariamente associado à responsabilidade moral para a ação é o **controle de direcionamento**" e não o controle regulativo (1998, p. 33, grifo do autor).

Diversos fatores podem minar o controle de direcionamento. Se o comportamento de uma pessoa for provocado por hipnose, lavagem cerebral, ou impulsos genuinamente irresistíveis, então essa pessoa não pode ser moralmente responsável por seu comportamento, visto que ela não o guia reflexivamente da maneira que é exigida para a responsabilização (FISCHER; RAVIZZA, 1998, p. 35). Mais especificamente, uma agente nas circunstâncias acima não parece ser responsável porque ela "não é sensível às razões – seu comportamento seria o mesmo, não importa quais razões houvesse." (1998, p. 37). Assim, Fischer e Ravizza caracterizam a posse de controle de direcionamento como (parcialmente) dependente da sensibilidade às razões. Mais especificamente, o controle de direcionamento depende de saber se o **mecanismo** psicológico do qual parte o comportamento da agente é sensível às razões (o controle regulador também exige que uma agente **possua** o mecanismo sobre o qual ela age. De acordo com Fischer e Ravizza, essa exigência realoca as condições históricas sob a responsabilidade. (§ 3.3.3).

O foco de Fischer e Ravizza nos mecanismos é motivado pelo seguinte raciocínio. Em um caso de Frankfurt, uma agente é responsável por uma ação mesmo que sua ação esteja assegurada por fatores externos. Mas a presença de tais fatores externos significa que a agente do caso de Frankfurt teria agido da mesma forma, não obstante as razões que tenha confrontado, o que sugere que a agente responsável em um cenário Frankfurt não é sensível a razões. Aqui temos um problema para a afirmação de Fischer e Ravizza de que o controle de direcionamento e, assim, a sensibilidade a razões, é necessária para a responsabilização.

A solução dos autores consiste em argumentar que, apesar de a **agente,** em um caso de Frankfurt, poder não ser sensível a razões, seu mecanismo, "o

processo que leva ao resultado relevante, isto é, à ação do agente", pode bem ser sensível a razões (1998, p. 38). Em outras palavras, o mecanismo psicológico genericamente-especificado da agente pode ter respondido (sob as condições contrafactuais) a considerações em favor de evitar a ação que a agente realmente executou (e que ela estava fadada a executar, independentemente das razões, visto que ela estava em um cenário do tipo Frankfurt).

Assim, Fischer e Ravizza chegam à seguinte conclusão provisória: "casos relativamente claros de responsabilidade moral", isto é, aqueles nos quais uma agente não é hipnotizada, etc., são distinguidos pelo fato de que "uma agente exibe controle de direcionamento de uma ação na medida em que o mecanismo em questão na ação é seu próprio mecanismo, sensível a razões" (1998, p. 39). Mas o **quão** sensível a razões o mecanismo de uma agente precisa ser para que ela tenha o tipo de controle sobre o seu comportamento associado à responsabilidade moral? Um mecanismo **fortemente** sensível a razões **reconheceria** e **responderia a qualquer** razão **suficiente** para agir de outro modo (1998, p. 41). (Na terminologia de Fischer e Ravizza, este mecanismo é fortemente "receptivo" e "reativo" a razões). No entanto, uma forte sensibilidade a razões pode não ser exigida para o controle de direcionamento, visto que muitas agentes intuitivamente sensíveis — isto é, muitas malfeitoras — deixam de atender às razões suficientes para agir de outro modo. Por outro lado, uma sensibilidade a razões *fraca* não é suficiente para o controle de direcionamento. Uma agente com um mecanismo fracamente sensível a razões responderá apropriadamente a **alguma** razão suficiente para agir de outro modo, mas o padrão de sensibilidade revelado no comportamento da agente pode ser muito arbitrário para que ela seja considerada como tendo o tipo de controle requerido para a responsabilização.

O padrão de sensibilidade a razões da pessoa possivelmente pareceria irregular de um modo relevante se, por exemplo, ela desistisse de comprar um ingresso para um jogo de futebol caso custasse mil reais, mas não se custasse dois mil reais. (FISCHER; RAVIZZA, 1998, p. 66).

Fischer and Ravizza decidem pela sensibilidade **moderada** a razões como o tipo mais pertinente ao controle de direcionamento (1998, p. 69-85). Um mecanismo psicológico que é moderadamente sensível às razões exibe regularidade com respeito à receptividade a razões: isto é, exibe "um **padrão compreensível** da (real e hipotética) receptividade a razões" (FISCHER; RAVIZZA, 1998, p. 71, grifo do autor). Tal padrão indicará que uma agente entende "como as razões se encaixam" e que, por exemplo, a "aceitação de uma razão como suficiente implica que uma

razão mais forte também deva ser suficiente" (FISCHER; RAVIZZA, 1998, p. 71). Adicionalmente, um padrão de receptividade regular a razões incluirá receptividade a uma gama de considerações **morais** (FISCHER; RAVIZZA, 1998, p. 77). Isso excluirá a atribuição de responsabilidade moral a agentes não morais. A esse respeito, confira Todd e Tognazzini (2008) para críticas à articulação desta condição por Fischer e Ravizza. Entretanto, um mecanismo moderadamente sensível pode ser apenas fracamente **reativo** às razões, na medida em que, como Fischer e Ravizza acrescentam, de modo um pouco misterioso, que "a reatividade é consistente", de tal forma "que se o mecanismo de uma agente reage a algum incentivo para... [agir de outro modo], isso mostra que o mecanismo **pode** reagir a **qualquer** incentivo para agir de outro modo." (1998, p. 73, grifo do autor).

A teoria de Fischer e Ravizza gerou muita atenção e críticas. Algumas críticas focam no contraste (já mencionado) entre as condições que os autores impõem sobre a receptividade às razões e as condições que impõem sobre a reatividade a razões (MCKENNA, 2005; MELE, 2006a; WATSON, 2001). Ademais, muitas pessoas ficaram insatisfeitas com a apresentação de Fischer e Ravizza de sua teoria em termos de poderes dos mecanismos em oposição a agentes. Isso levou algumas autoras e alguns autores a desenvolverem teorias da sensibilidade a razões baseadas em agentes, que se remetem aos problemas que levaram Fischer e Ravizza à sua abordagem baseada em mecanismos (BRINK; NELKIN, 2013; MCKENNA, 2013; SARTORIO, 2016).

3. Debates Contemporâneos

3.1. As "Faces" da Responsabilidade

3.1.1. Atributabilidade *versus* Imputabilidade

Será que as nossas práticas de responsabilização acomodam formas distintas de responsabilidade moral? Será que existem diferentes sentidos nos quais as pessoas podem ser moralmente responsáveis por seus comportamentos? O interesse contemporâneo nessas possibilidades tem raíz no debate entre Susan Wolf e Gary Watson. Entre outras coisas, o importante livro de Wolf, *Freedom Within*

Reason (1990), oferece uma discussão crítica das teorias sobre responsabilização do "Eu Verdadeiro". Com efeito, uma pessoa é responsável por um comportamento que é atribuível ao seu Eu verdadeiro e

> o comportamento de uma agente é atribuível ao Eu verdadeiro da agente... se ela tem a liberdade (ou é capaz) de governar seu comportamento com base em sua vontade como também capaz de governar sua vontade com base em seu sistema de avaliação. (WOLF, 1990, p. 33).

A ideia básica é que uma agente responsável não é apenas movida por seus desejos mais fortes, mas também, de algum modo, aprova ou apoia os desejos que a movem porque eles são governados por seus valores ou porque eles são endossados por algum desejo de ordem superior. O exemplo central de Wolf da tese do Eu verdadeiro é de Watson (1975). Em um importante e intimamente relacionado artigo anterior, Wolf (1987) caracteriza Watson (1975), Harry Frankfurt (1971), e Charles Taylor (1976) como oferecendo "uma imagem do eu profundo". Para saber mais sobre as abordagens do Eu-verdadeiro/Eu-profundo, veja §3.3.3. Já para uma apresentação recente de uma abordagem de um Eu-verdadeiro, confira Chandra Sripada (2016).

De acordo com Wolf, um ponto a favor das teorias do Eu verdadeiro é que elas explicam por que pessoas agindo sobre a influência de hipnose ou desejos compulsivos não são frequentemente responsabilizadas (1990, p. 33). Visto que essas agentes são tipicamente incapazes, sob essas condições, de governar seu comportamento com base em seu sistema valorativo, elas são alienadas das suas ações de um modo que exclui sua responsabilidade. Mas, para Wolf, um ponto contra as teses do Eu verdadeiro é que elas tendem a ser silenciosas com relação ao tópico de como agentes passaram a possuir o Eu que têm. O Eu verdadeiro de uma agente poderia, por exemplo, ser um produto de uma experiência traumática, e Wolf argumenta que isso nos daria razões para questionar a "responsabilidade da agente **por** seu Eu verdadeiro" e, assim, sua responsabilidade pelo comportamento atual que surge desse Eu (1990, p. 37, grifo do autor). Para uma importante teoria de uma agente com tal formação, veja o exemplo ficcional de Jojo, dado por Wolf (1987), e Watson 2004 para a discussão relacionada sobre o assassino condenado Robert Alton Harris. Para a discussão de Jojo deste verbete, confira § 3.2.1, e para

uma discussão geral sobre a relevância da história pessoal para a responsabilidade atual, consulte § 3.3.3.

Wolf sugere que quando o "Eu verdadeiro" de uma pessoa é o produto de graves traumas de infância (ou fatores relacionados), esta pessoa é potencialmente responsável por seu comportamento somente em um sentido superficial, que meramente atribui más ações ao seu "Eu verdadeiro" (1990, p. 37-40). Entretanto, Wolf argumenta que imputações de responsabilidade moral vão muito mais além do que tais atribuições podem alcançar:

> Quando [...] consideramos um indivíduo digno de culpa ou elogio, não estamos apenas julgando a qualidade moral do evento com o qual o indivíduo está tão intimamente associado; estamos julgando a qualidade moral do próprio indivíduo de uma maneira mais focada, não-instrumental e aparentemente mais séria (1990, p. 41).

Essa forma mais profunda de avaliação, em termos de "responsabilização profunda" (WOLF, 1990, p. 41), requer mais do que a agente ser "capaz de formar suas ações com base em seus valores"; também exige que "ela seja capaz de formar seus valores com base no que é Verdadeiro e Bom" (WOLF, 1990, p. 75). Esta última habilidade será debilitada ou ausente em uma agente cujo Eu verdadeiro é o produto de pressões (tal qual uma infância traumática) que tenham distorcido sua visão moral (para a relevância da visão moral, ou "competência moral" para a responsabilidade, ver §3.2).

Em *Two Faces of Responsibility*, de 1996 (2004), Gary Watson responde a Wolf, concordando com ela que algumas teorias da responsabilização – isto é, **teses de autoexposição** – uma frase que Watson empresta de Benson (1987) – concentram-se de modo estreito no problema de saber se o comportamento é atribuível a uma agente. Mas Watson nega que tais atribuições sejam uma mera forma superficial de avaliação da responsabilidade. Afinal, o comportamento que é atribuível a uma agente, no sentido, por exemplo, de partir do seu sistema valorativo, muitas vezes expõe algo interpessoal e moralmente significativo sobre a "orientação avaliativa fundamental" da agente (WATSON, 2004, p. 271), de modo que as imputações de responsabilidade neste sentido de **responsabilidade como imputabilidade** são "centrais à vida ética e à apreciação ética." (WATSON, 2004, p. 263).

Porém, Watson concorda com Wolf que a história da responsabilização citada acima está incompleta: há mais a ser dito sobre a responsabilidade do que atribuir ações às agentes. Ademais, **consideramos agentes como responsáveis** por seu comportamento, o que "não é apenas uma questão da relação entre um indivíduo e seu comportamento" (WATSON, 2004, p. 262). Quando consideramos alguém como responsável, também "demandamos determinada conduta e reagimos adversamente à incapacidade das pessoas de se conformarem a tais demandas" (WATSON, 2004, p. 262). As demandas morais, e o potencial para o tratamento adverso associado à consideração das outras pessoas como responsáveis, são parte das nossas práticas de incumbência (como opostas à atribuibilidade), e essas características da incumbência levantam questões de justiça que não são levantadas no contexto de determinar se o comportamento é atribuível a uma agente (WATSON, 2004, p. 273). Portanto, certas condições podem ser aplicadas à incumbência que não são aplicáveis à atribuibilidade: por exemplo, talvez a "culpa por incumbência" devesse ser, como Wolf sugere, moderada no caso de uma agente cujas "circunstâncias esquálidas tornaram esmagadoramente difícil desenvolver o respeito pelos padrões para os quais a consideraríamos incumbida" (WATSON, 2004, p. 281).

Portanto, existem duas formas ou "faces" da teoria de Watson sobre a responsabilidade. Há a responsabilidade como atribuibilidade, e quando uma agente satisfaz as condições dessa forma de responsabilidade, o comportamento lhe é devidamente atribuído de modo a refletir as características moralmente importantes do seu Eu, suas virtudes e vícios, por exemplo. Mas há também a responsabilidade como incumbência, e quando uma agente satisfaz as condições dessa forma de responsabilidade, as quais requerem mais do que a correta atribuição do comportamento, ela está aberta a ser incumbida por um comportamento de maneira que caracteriza predominantemente a culpa moral.

3.1.2. Atribuicionismo

Tem sido comum descrever as teses de diversas autoras e diversos autores (com graus variados de precisão) como instâncias do "atribuicionismo", confira Neil Levy (2005) para o primeiro uso do termo, incluindo Robert Adams (1985), Nomy Arpaly (2003), Pamela Hieronymi (2004), TM Scanlon (1998, 2008), George Sher (2006a, 2006b, 2009), Angela Smith (2005, 2008) e Matthew Talbert (2012, 2013). Atribuicionistas consideram a avaliação da responsabilidade moral como sendo

principalmente preocupada com o fato de saber se uma ação (ou omissão, traços de caráter, ou crença) é atribuível a uma agente para fins de avaliação moral, o que usualmente significa que a ação (ou omissão, etc.) reflete as "atitudes sensíveis ao julgamento" da agente (SCANLON, 1998), os seus "julgamentos avaliativos" (SMITH, 2005), ou, mais geralmente, a sua "personalidade moral" (HIERONYMI, 2008).

O atribuicionismo se assemelha às teses de autoexposição mencionadas por Watson (*vide* a subseção anterior) na medida em que ambas focam no modo em que a agente responsável expõe características que são significativas interpessoal e moralmente em relação ao "Eu". Entretanto, seria um engano concluir que as posições do atribuicionismo contemporâneo estão interessadas somente em especificar as condições às quais Watson chama de responsabilidade como atribuibilidade. De fato, atribuicionistas tipicamente se consideram como oferecendo condições para sustentar as agentes como responsáveis no sentido de **incumbência** colocado por Watson (*vide* a subseção prévia para a distinção entre **incumbência** e **atribuibilidade**).

De acordo com o atribuicionismo, o preenchimento das condições de atribuibilidade é suficiente para considerar as agentes como incumbidas por seu comportamento. Isto significa que o atribuicionismo rejeita condições de responsabilização moral que perdoariam as agentes se seu caráter fosse moldado sob condições adversas (SCANLON, 1998, p. 278-285), ou se aquilo pelo qual a agente é culpada não estava sob seu controle (SHER, 2006b, 2009; SMITH, 2005), ou se não se pode esperar que a agente reconheça o *status* moral de seu comportamento (SCANLON, 1998, p. 287-290; TALBERT, 2012). Atribuicionistas rejeitam essas condições sobre a responsabilização porque o comportamento significante moral e interpessoalmente é atribuível às agentes que não as preenchem, e tais atribuições são tomadas como sendo suficientes para que uma agente esteja aberta às reações envolvidas em considerar agentes como incumbidas por seu comportamento. Atribuicionistas também argumentam que a culpa pode ser proficuamente entendida como uma forma de protesto moral (HIERONYMI, 2001; SMITH, 2013; TALBERT, 2012), visto que parte do apelo desse movimento é que protestos morais podem ser legítimos nos casos em que as condições acima não são encontradas.

Diversas objeções foram feitas ao atribuicionismo. Algumas autoras e alguns autores argumentam que atribuicionistas estão errados e/ou erradas em rejeitar as condições sobre a responsabilização mencionadas no último parágrafo (LEVY, 2005, 2011; SHOEMAKER, 2011, 2015a; WATSON, 2011). Também foi argumentado que a posição atribuicionista de culpa é muito próxima da mera

avaliação negativa (LEVY, 2005; WALLACE, 1996, p. 80-81; Watson, 2002). Além disso, Scanlon (2008) tem sido criticada por não ter considerado emoções negativas como ressentimento, como centrais para o fenômeno da culpa (WALLACE, 2011; WOLF, 2011). Uma crítica similar se aplicaria a Sher (2006a).

3.1.3. Responsividade

Ao desenvolver a distinção entre atribuibilidade e incumbência (§3.1.1), David Shoemaker (2011, 2015a) introduziu uma terceira forma de responsabilidade: responsividade. Na visão de Shoemaker, as avaliações de responsabilidade atributiva respondem aos fatos sobre o caráter de uma agente, a responsabilidade como incumbência refere-se ao grau de consideração de uma agente para com outrem, e a responsabilidade enquanto responsividade atende aos julgamentos de avaliação dela. Entretanto, A. Smith (2015) e Hieronymi (2008, 2014) usam "responsividade" para se referirem a uma visão mais parecida com a perspectiva atribuicionista descrita na subseção anterior, e Pereboom (2014) usou o termo para indicar uma forma da responsabilidade conveniente aos céticos acerca responsabilização moral.

3.2. Competência Moral

3.2.1. A Condição da Competência Moral sobre Responsabilidade

A possibilidade de que a competência moral, a habilidade de reconhecer e responder a considerações morais, seja uma condição sobre a responsabilidade moral foi sugerida em vários pontos acima (§2.2.1, §2.2.2, §2.3, §3.1.1, §3.1.2). A história fictícia de Susan Wolf (1987) sobre "JoJo" é uma das ilustrações mais conhecidas desta proposta. JoJo foi educado por um vil ditador, e, como resultado, tornou-se o mesmo tipo de tirano sádico que seu pai. Como adulto, JoJo é feliz sendo o tipo de pessoa que é, movido precisamente pelos desejos que ele quer que o movam (aprisionar, torturar, e executar os seus subordinados). Assim, JoJo preenche importantes condições da responsabilização (§3.1.1, §3.3.3), contudo, Wolf argumenta que pode ser injusto considerá-lo como responsável por seu mau comportamento.

A educação de JoJo desempenha um importante papel no argumento de Wolf, mas somente porque torna JoJo incapaz de apreciar totalmente as malevolências de seu comportamento. Assim, é a competência moral debilitada de JoJo que o torna realmente desculpável, e conclusões similares de não responsabilização deveriam ser tiradas sobre todas as pessoas que achamos que "não poderiam deixar de estar enganadas sobre seus [maus] valores" se a posse destes valores debilitam sua habilidade de diferenciar o certo do errado (WOLF, 1987, p. 57).

Muitas outras e muitos outros se juntam a Wolf ao argumentar que a competência moral debilitada, talvez por causa da formação ou fatores ambientais, elimina a responsabilidade moral de uma pessoa (BENSON, 2001; DORIS; MURPHY, 2007; FISCHER; RAVIZZA, 1998; FRICKER 2010; LEVY 2003; RUSSELL, 1995, 2004; WALLACE, 1996; WATSON, 2004). Parte do que motiva essa conclusão é o pensamento de que pode não ser razoável esperar que agentes moralmente debilitadas evitem o comportamento errado, e que, portanto, é injusto expor essas agentes aos danos da culpa moral por causa de seu mau comportamento. Para um desenvolvimento detalhado do requerimento da competência moral sobre a responsabilidade em termos de consideração de justiça, confira Jay Wallace (1996) e, também, Erin Kelly (2013), Neil Levy (2009), e Gary Watson, de 1987 (2004). Para a rejeição da afirmação de que a culpa moral é injusta nos casos de agentes moralmente debilitadas, ver os diversos defensores do atribuicionismo mencionados em §3.1.2, particularmente Hieronymi (2004), Scanlon (1998) e Talbert (2012).

A condição de competência moral para a responsabilização pode também ser motivada pela sugestão de que agentes debilitadas não são capazes de cometer erros que tenham o tipo de significância moral para a qual a culpa seria uma resposta apropriada. A ideia básica aqui é que, apesar das agentes moralmente debilitadas poderem deixar de mostrar o respeito apropriado a outras pessoas, estas falhas não necessariamente constituem o tipo de desdém às normas morais que fundamentam a culpa (Watson 2004, p. 234). Em outras palavras, **uma falta de respeito** para com as outras pessoas nem sempre é uma instância de culpa fundamentada pelo **desrespeito** às outras pessoas, visto que o **desrespeito** (mas não qualquer falta de respeito) requer a habilidade de compreender as normas que se viola (LEVY, 2007; SHOEMAKER, 2011).

3.2.2. Teorias Conversacionais sobre a Responsabilidade

As considerações sobre a competência moral desempenham um importante papel nas recentes tendências das teorias **conversacionais** da responsabilidade, que concebem os elementos de nossas práticas de responsabilização como movimentos moralmente expressivos em uma conversa moral em andamento. A ideia aqui é que, para participar proficuamente, e em sua totalidade, de uma conversa, alguém deve ter algum grau de competência na linguagem moral dessa conversação.

Diversas versões proeminentes da abordagem conversacional desenvolvem a sugestão de P. F. Strawson (§ 2.2.1) de que as atitudes reativas negativas envolvidas na culpa são expressões de uma **demanda** por consideração moral de outras agentes. Gary Watson, em estudo de 1987, argumenta que a demanda "presume", como sua condição, a inteligibilidade de expressá-la, o "entendimento por parte do objeto da demanda" (2004, p. 230). Portanto, visto que "as atitudes reativas são formas incipientes de comunicação", elas são inteligivelmente expressas "somente na assunção de que a outra pessoa pode compreender a mensagem" e, visto que a mensagem é moral, "culpar e louvar as pessoas que têm um entendimento moral diminuto se desvirtua de sua 'finalidade'", ao menos num certo sentido (2004, p. 230). Consulte a obra de Watson, de 2011, para uma modificação da proposta. R. Jay Wallace argumenta que, dado que as práticas sinceras de responsabilidade são inerentes às relações morais que são

> definidas pela troca bem-sucedida de crítica e justificação moral [...] será razoável considerar como incumbida apenas uma pessoa que seja ao menos candidata a esse tipo de troca de crítica e justificação (1996, p. 164).

Conversation and Responsibility (2012), de Michael McKenna, oferece a análise conversacional mais desenvolvida sobre a responsabilidade. Para Mckenna, a "troca da responsabilidade moral" ocorre em estágios: uma "contribuição moral" inicial de um comportamento moralmente substancial; a "consideração moral" da culpa, por exemplo, que responde à contribuição moral; a "teoria moral", na qual o primeiro colaborador responde pela consideração moral com, por exemplo, um pedido de desculpas; e assim por diante (2012, p. 89). Assim como Wallace e Watson, Mckenna ressalta o modo no qual uma agente moralmente debilitada

encontrará dificuldade para “apreciar os desafios que lhe são colocados por aquelas pessoas que a consideram moralmente responsável”, mas o autor também argumenta que uma agente apropriadamente debilitada não pode sequer fazer o primeiro movimento em uma conversa moral (2012, p. 78). Portanto, a responsabilidade da agente moralmente debilitada é colocada em questão não apenas porque ela é incapaz de responder apropriadamente às demandas morais, mas também porque “é incapaz de agir a partir de uma vontade com a qualidade moral que poderia ser candidata à avaliação do ponto de vista da responsabilização.” (McKenna 2012, p. 78). Esse ponto está relacionado com a disputa entre Neil Levy e David Shoemaker, mencionada na subseção anterior, de que debilidades na competência moral podem deixar uma agente incapaz de nutrir e expressar o tipo de má vontade ou ausência de consideração às quais a culpa corresponde. Por outro lado, Watson (2011) parece conceder que uma debilidade moral significativa seja compatível com a habilidade de realizar transgressões relevantes para a culpa, mesmo se tal debilidade eliminar a incumbência moral da transgressora por suas ações.

Veja a discussão de Shoemaker do tipo de raiva moral envolvida em considerar as outras pessoas como incumbidas por seu comportamento (2015a, p. 87-117) para outra concepção importante de responsabilidade em termos conversacionais amplos. Para defesas adicionais e articulações da abordagem conversacional à responsabilização, ver Stephen Darwall (2006), Miranda Fricker (2016), e Colleen Macnamara (2015).

3.2.3. Psicopatia

Debilidades na competência moral vêm em graus. O caso de JoJo, de Susan Wolf, (§ 3.2.1) apresenta debilidades localizadas na capacidade de reconhecer e responder a considerações morais, mas não é claro que ele seja imune a considerações morais. Entretanto, no extremo final do espectro, encontramos figuras mais global e profundamente debilitadas, tais como o psicopata. No tratamento filosófico, o psicopata é apresentado tipicamente como um agente que, apesar de reter outras capacidades psicológicas, é inteiramente ou tanto quanto possível, incapaz de responder apropriadamente às considerações morais (essa é uma construção filosófica, visto que a psicopatia na vida real admite vários graus de debilidade, correspondendo a pontuações mais altas ou mais baixas nas medições de diagnóstico).

Uma questão interessante é a de saber se a incapacidade do psicopata de responder apropriadamente aos incentivos morais seja primariamente o resultado de uma falha motivacional, em vez de cognitiva: será que o psicopata sabe o que a moralidade requer, mas simplesmente não se importa? Se uma resposta positiva for dada a esta última questão (FISCHER; RAVIZZA, 1998, P. 76-81; NICHOLS, 2002), então parece provável que o psicopata seria responsável por ao menos parte de seu mau comportamento. E algumas pessoas argumentam que mesmo se a psicopatia for principalmente uma debilidade cognitiva, ainda poder ser o caso de que psicopatas possuam capacidades suficientes para distinguir o certo e o errado, ou que possam possuir capacidades suficientemente relacionadas, para serem considerados como responsáveis, ao menos até certo ponto e de certos modos (GLANNON, 1997; GREENSPAN, 2003; MAIBOM, 2008; SHOEMAKER, 2014; VARGAS; NICHOLS, 2007). Por outro lado, muitos acreditam que a capacidade dos psicopatas de compreender considerações morais seja superficial demais para sustentar a responsabilização (KENNETT, 2019; LEVY, 2007; NELKIN, 2015; WALLACE, 1996, p. 177-78; WATSON, 2011). Para a afirmação de que a deficiência relevante é de conhecimento moral ao invés de capacidade moral, confira Mason (2017). Ademais, outras pessoas ainda argumentaram que mesmo aqueles que estão inteiramente debilitados em seu entendimento moral estão abertos à culpa na medida em que possuem ampla competência racional (SCANLON, 1998, p. 287-290; Talbert, 2014). No entanto, a posse dessas amplas competências por parte do psicopata foi colocada em questão por Fine e Kennett (2004), Greenspan (2003) e Litton (2010).

3.3. Ceticismo e tópicos relacionados

Esta seção introduz o ceticismo contemporâneo acerca da responsabilidade moral por meio de discussões de diversos tópicos que têm ampla relevância para o pensamento sobre a responsabilidade.

Se a responsabilidade moral exige o livre-arbítrio, e o livre-arbítrio envolve o acesso a alternativas de uma maneira que não é compatível com o determinismo, então, se seguiria da verdade do determinismo que ninguém nunca é moralmente responsável. O raciocínio acima e a conclusão cética a que chega são endossados pela perspectiva do **determinismo rígido** sobre o livre-arbítrio e a responsabilidade. Essa perspectiva foi defendida, historicamente, por Spinoza e d'Holbach, entre

outras pessoas, e por Ted Honderich (2002), mais recentemente. Mas, dado que o determinismo pode ser falso, os céticos contemporâneos sobre a responsabilidade moral perseguem com mais frequência uma linha de argumentação **incompatibilista rígida**, segundo a qual o tipo de livre-arbítrio requerido para a responsabilidade moral baseada no merecimento (em oposição à teoria prospectiva, *vide* § 2.1) é incompatível com a **verdade ou falsidade** do determinismo (PEREBOOM, 2001, 2014). As posições céticas discutidas abaixo são geralmente deste tipo: as conclusões céticas que advogam não dependem da verdade do determinismo.

3.3.1. Sorte Moral

De acordo com Thomas Nagel, uma pessoa é um sujeito de sorte moral se fatores que não estão sob seu controle afetam a avaliação moral sob a qual ela está sujeita. Nesse sentido, confira o estudo de Nagel, de 1976 (1979), e, ainda, o trabalho de Williams (1981), também de 1976, além de o verbete sobre **sorte moral**.

A sorte moral realmente existe? Mais especificamente, a sorte moral pode afetar a responsabilização? Considere um pretenso assassino que atira em seu alvo, no intuito de matar, mas falha em fazê-lo somente porque sua bala é desviada por um pássaro que interfere em sua trajetória. Parece que este pretenso assassino teve um **resultado moral fortuito** (isto é, uma sorte moral fortuita no resultado de seu comportamento). Por causa de fatores além do seu controle, o histórico moral do pretenso assassino é melhor do que poderia ter sido: particularmente, ele não é um assassino e não é moralmente responsável por ter causado a morte de alguém. Alguém poderia pensar, além disso, que o pretenso assassino é menos culpado do que um assassino bem-sucedido com o qual é idêntico em todos os outros aspectos, e que a razão para isso é apenas que o assassino bem-sucedido matou alguém intencionalmente, enquanto o pretenso assassino não (como resultado da sorte moral fortuita). Para uma recente e importante defesa da sorte moral, consulte os trabalhos de Hanna (2014) e Hartman (2017).

Por outro lado, alguém poderia pensar que se os dois assassinos mencionados acima são idênticos em termos de seus valores, objetivos, intenções e motivações, então a adição de um pouco de sorte na história do assassino malsucedido não pode fundamentar um profundo contraste entre esses dois agentes em termos de responsabilidade moral. Um modo de sustentar esta posição é argumentar que a responsabilidade moral é exclusivamente uma função das características inerentes

dos agentes, tais quais seus motivos e intenções. Para alguns dos principais argumentos contra a sorte moral, confira Khoury, 2018) e Enoch e Marmor (2007). É claro que o assassino bem-sucedido é responsável por **alguma coisa** (matar uma pessoa) pela qual o assassino malsucedido não o é, mas seria possível argumentar que ambos são moralmente responsáveis, e presumivelmente culpáveis, no **mesmo grau,** na medida em que foi verdadeiro que ambos intentaram matar e que eles o fizeram pelas mesmas razões e com o mesmo grau de comprometimento em provocar aquele resultado (*vide* M. Zimmerman, 2002, 2015).

Mas agora considere um pretenso assassino diferente, que nem mesmo tenta matar alguém, mas somente porque suas circunstâncias não favoreceram essa opção. Este pretenso assassino **deseja matar** sob circunstâncias favoráveis (e, portanto, ele parece ter tido sorte moral fortuita **circunstancial**, já que não se encontrava nessas circunstâncias favoráveis). Talvez o grau de responsabilidade atribuído ao assassino bem-sucedido e ao assassino malsucedido descrito acima não dependa tanto do fato de que ambos tentaram matar quanto do fato de que ambos desejavam matar; neste caso o pretenso assassino que acabamos de introduzir pode compartilhar do grau de responsabilidade dos assassinos anteriores, visto que ele compartilha de sua vontade de matar. Mas uma teoria que foca em como as agentes desejariam agir sob circunstâncias contrafactuais provavelmente geraria conclusões contra-intuitivas sobre a responsabilização, visto que muitas agentes que são tipicamente julgadas inocentes poderiam intencionalmente executar ações terríveis sob as circunstâncias adequadas. M. Zimmerman (2002, 2015) não se esquiva dessa consequência, mas as críticas de seus esforços para rejeitar a sorte moral, veja os estudos de Hanna (2014) e Hartman (2017), os quais têm dado muita importância a ela. Peels (2015) possui uma posição relacionada à de Zimmerman, mas que pode evitar a consequência contra-intuitiva que acabamos de mencionar.

Outra abordagem da sorte moral sustenta que esta é prejudicial à responsabilidade moral de um modo que geralmente elimina as atribuições de responsabilidade. Para ver a motivação dessa posição cética, considere a sorte moral **constitutiva**: isto é, sorte em como alguém é constituída em termos de "inclinações, capacidades, e temperamento" que encontra em si mesma (NAGEL, 1979, p. 28). Fatos sobre as inclinações, capacidades e temperamento de uma pessoa explicam muito, senão tudo, sobre o seu comportamento e, se os fatos que explicam por que alguém age tal como age são um resultado de boa ou má sorte, então talvez seja injusto responsabilizá-la por tal comportamento. Como Nagel

aponta, uma vez que a paisagem completa dos vários tipos de sorte surge à vista, "a área da agência genuína" parece se reduzir à nada, visto que as nossas ações e suas consequências "resultam de uma combinação de fatores e influências, anteriores e posteriores à ação, que não estão dentro do controle da agente." (1979, p. 35). Se isso for correto, então talvez

> nada resta do que possa ser atribuído ao "EU" responsável, e ficamos com nada além de uma [...] sequência de eventos, que podem ser lastimados ou celebrados, mas não culpados ou elogiados (NAGEL, 1979, p. 37).

Apesar das citações acima, Nagel não adere totalmente a uma conclusão cética acerca da responsabilidade com base na sorte moral, mas outras e outros o fizeram, mais notavelmente, Neil Levy (2011). De acordo com a "visão rígida da sorte" de Levy, a natureza abrangente da sorte moral significa que "não existem diferenças que acarretam merecimento entre os agentes morais." (2011, p. 10). É claro que existem diferenças entre agentes em termos do caráter, das boas ou más ações e dos resultados que elas produzem, mas o ponto de Levy é que, dada a influência da sorte na geração dessas diferenças, elas não fornecem uma base sólida para o tratamento diferenciado das pessoas em termos de elogio ou culpa moral. Para uma teoria compatibilista, que leva a uma variedade de pessimismos, embora não ao ceticismo, com base nas questões sobre sorte moral já descritas, consulte o estudo de Russell (2017).

3.3.2. A Responsabilidade Última

Outro importante argumento cético, relacionado às observações da subseção anterior acerca da sorte moral constitutiva, é o Argumento Básico de Galen Strawson, que conclui que "não podemos ser verdadeiramente ou ultimamente responsáveis pelas nossas ações" (1994, p. 5). Visto que o argumento visa a responsabilidade moral "última", ele não necessariamente exclui outras formas de responsabilidade, tal como a responsabilidade prospectiva (§ 2.1) e, de acordo com algumas interpretações, a responsabilidade como atribuibilidade (§ 3.1.1). O argumento começa por notar que uma agente faz as escolhas que faz por causa de certos

fatos sobre como ela é: por exemplo, os fatos sobre o que lhe parece digno de escolha. Mas, se isso for verdadeiro, então, a fim de ser responsável por suas escolhas subsequentes, talvez uma agente também precise ser responsável pelos fatos que lhe parecem ser dignos de escolha. Mas como alguém pode ser responsável por esses fatos anteriores sobre si mesmo? Isso não exigiria uma escolha anterior por parte da agente, algo que resultaria nas suas disposições presentes para ver certos fins e meios como dignas de escolha? Mas essa escolha anterior seria em si mesmo algo pela qual a agente é responsável somente se a agente é também responsável pelo fato de que *essa* escolha anterior lhe pareceria preferível. E agora devemos explicar como a agente pode ser responsável por esse fato anterior adicional sobre si mesma, o qual irá requerer postular outra escolha pela agente, e a responsabilidade pela escolha que também terá de ser assegurada, o que requereria explicar por que isso lhe pareceu preferível, e assim por diante. Um regresso surge, e Strawson afirma que ele não pode ser parado, exceto pela postulação de um ato inicial de autocriação por parte da agente responsável (1994, p. 5, p. 15). Somente agentes que criam a si mesmas poderiam ser totalmente responsáveis por suas próprias tendências de exercitar seus poderes de escolha tal qual o fazem, mas a autocriação é impossível, então, ninguém é verdadeira ou ultimamente moralmente responsável por seu comportamento.

Várias réplicas a esse argumento (e ao argumento da sorte moral constitutiva) são possíveis. Alguém poderia simplesmente negar que **como** uma pessoa veio a ser do modo como ela é seja uma questão relevante para a responsabilização: talvez tudo que precisemos saber, a fim de julgar a responsabilidade atual, sejam os fatos sobre sua constituição presente e sobre como esta constituição está relacionada a seu comportamento atual (*vide* **atribuicionismo** § 3.1.2 e a discussão das teorias não históricas na próxima subseção). Alternativamente, alguém poderia pensar que, apesar da história pessoal importar para responsabilização moral, o argumento de Strawson é muito exigente, requerendo **demasiado** controle histórico sobre a constituição de uma pessoa (*vide* Fischer, 2006). Para uma réplica, consulte o estudo de Levy (2011, p. 5). Talvez o necessário não seja literalmente a autocriação, mas simplesmente uma habilidade de ordenar mudanças em si mesmo de modo a adquirir responsabilidade pelo Eu que resulte dessas mudanças (Clarke, 2005). Uma imagem neste sentido pode ser encontrada na sugestão de Aristóteles de que uma pessoa pode ser responsabilizada por ser negligente, se seu estado atual de negligência for o resultado de uma escolha anterior que essa pessoa tenha feito ([Arist.] *Ethica Nichomachea*). Confira, também, Michele Moody-Adams (1990).

Nessa linha aristotélica, Robert Kane oferece uma explicação incompatibilista detalhada de como podemos assegurar a responsabilidade última por nossas ações (1996, 2007). Na visão de Kane, para uma agente

> ser *ultimamente* responsável por uma escolha, a agente deve ser pelo menos em parte responsável em virtude de escolhas ou ações voluntariamente realizadas no passado por ter o caráter e os motivos que ela agora têm (2007, p. 14, grifo do autor).

Pode parecer que essa visão esteja sujeita às preocupações acerca do regresso apresentadas acima pelo argumento de Galen Strawson. Mas Kane pensa que um regresso é evitado em casos nos quais as escolhas formadoras de caráter de uma agente são indeterminadas. Visto que essas escolhas indeterminadas não terão causas **suficientes**, não há causa relevante anterior pela qual a agente deva ser responsável, e, então, não há o problema do regresso (Kane, 2007, p. 15-16). Para uma crítica a Kane, em relação a esse ponto, consulte Pereboom (2001, p. 47-50).

As escolhas formadoras potenciais do caráter que ocorrem "quando estamos divididos ou estamos divididas entre visões concorrentes do que devemos fazer ou nos tornar." (2007, p. 26) é particularmente interessante para Kane. Em tais casos, se uma pessoa encontra razões em favor de qualquer escolha que possa fazer, e a escolha pela qual opta é indeterminada, então, qualquer escolha que ela faça terá sido escolhida por suas próprias razões. De acordo com Kane, quando uma agente faz esse tipo de escolha, molda seu caráter e, visto que sua escolha não é determinada por fatores causais anteriores, a agente é responsável pela escolha, pelo caráter que molda e pelas escolhas determinadas pelo caráter que a própria agente irá constituir no futuro.

A abordagem de Kane é uma instância importante daquelas teorias incompatibilistas que tentam explicar como o livre-arbítrio, apesar de exigir o indeterminismo, poderia ser claramente acomodado no mundo natural tal como o conhecemos (BALANGUER, 2010; EKSTROM, 2000; FRANKLIN, 2018). Isto se opõe às teorias da causação da agente acerca do livre-arbítrio (CHISHOLM, 1964), e que envolvem um tipo de poder causal que é menos facilmente naturalizado (O'CONNOR, 2000). Entretanto, muitas pessoas argumentam que qualquer teoria como a de Kane, a qual insere um elo indeterminista à cadeia causal que leva à ação, na verdade reduz o controle da agente sobre uma ação ou ao menos não

deixa claro o porquê de tal inserção aumentar o controle da agência sobre as ações quando comparado a uma história determinística da ação (HOBART, 1934; LEVY, 2011, p. 41-83; PEREBOOM, 2014, p. 31-49; VAN INWAGEN, 1983, p. 126-52; WATSON, 1999).

3.3.3. História Pessoal e Manipulação

As teorias, como as de Neil Levy (2011) e de Galen Strawson (1994), descritas nas duas subseções precedentes, assumem que os fatos sobre o modo como uma pessoa veio a ser como ela é, são relevantes para determinar sua responsabilidade atual. Mas visões não históricas, como o atribuicionismo (§ 3.1.2) e as visões que Susan Wolf chama de teorias do "Eu Verdadeiro" (§ 3.1.1), rejeitam essa opinião. As teorias do "Eu verdadeiro" são, por vezes, referidas como teorias "estruturais" ou "hierárquicas", e John M. Fischer e Mark Ravizza (1998, p. 184-187) as denominaram teorias de "entrosamento". Seja qual for o nome, a ideia básica é que uma agente é moralmente responsável na medida em que sua vontade tem o tipo certo de estrutura: em particular, precisa haver um entrosamento ou encaixe entre os desejos que realmente movem a agente e os seus valores, ou entre os desejos que a movem e seus desejos de ordem superior, sendo estes últimos aquelas preferências reflexivas da agente, sobre quais desejos devem movê-la (*vide* DWORKIN, 1970; FRANKFURT, 1971,1987; NEELY, 1974; WATSON, 1975).

A comparação de Harry Frankfurt entre um viciado voluntário e um viciado involuntário ilustra características importantes de sua versão da abordagem estrutural da responsabilização. Ambos os viciados citados por Frankfurt têm desejos em usar a droga na qual são viciados e a natureza de seus vícios é tal que ambos desejam ultimamente agir para satisfazer seus desejos viciosos de primeira ordem. Mas suponha que ambos os viciados sejam capazes de formar perspectivas de ordem superior sobre seus desejos de primeira ordem, e suponha que eles considerem **diferentes** perspectivas de ordem superior. O viciado voluntário endossa e se identifica com o seu desejo vicioso. O viciado involuntário, por outro lado, repudia seu desejo vicioso a tal ponto que, quando termina por ser vencido por esse desejo, Frankfurt diz que ele é "desamparadamente violado pelos seus próprios desejos" (1971, p. 12). O viciado voluntário tem um tipo de liberdade da qual o involuntário carece: ambos podem estar fadados a usar a droga na qual são viciados, mas na medida em que o viciado voluntário é movido pelo desejo que endossa, ele age livremente

de um modo que o viciado involuntário não age (Frankfurt 1971, p. 19). Uma conclusão relacionada com a responsabilidade pode ser extraída: talvez o desejo do viciado involuntário seja alheio a si de tal modo que sua responsabilidade por agir com base nele é colocada em questão (*vide* SRIPADA, 2017).

Uma objeção à visão de Frankfurt se dá da seguinte maneira: sua teoria parece assumir que os desejos de ordem superior dos viciados tem autoridade para falar por eles, os desejos revelam (ou constituem) o "Eu verdadeiro" das agentes, para usar a linguagem de Wolf (1990). Mas se os desejos de ordem superior são invocados a partir de uma preocupação de que os desejos de primeira ordem da agente não se originam do seu "Eu verdadeiro", então por que a mesma preocupação não se repetiria também em relação aos desejos de ordem superior? Em outras palavras, quando ascendemos pelas ordens dos desejos, por que parar em qualquer ponto particular? Por que não pensar que apelar a uma ordem ainda mais elevada é sempre necessário para revelar as posições de uma agente? Para uma objeção nesses termos, a qual parcialmente motivou Watson, em sua articulação de uma abordagem estrutural, a focar na questão de saber se os desejos de uma agente estão em conformidade com seus **valores**, em vez de seus desejos de ordem superior, consulte Watson (1975).

E mesmo se concordássemos com Frankfurt (ou Watson) sobre os elementos estruturais requeridos para a responsabilização, alguém ainda poderia se perguntar: como a vontade de uma agente veio a ter sua estrutura particular? Assim, uma objeção importante à tese de Frankfurt aponta que a estrutura relevante poderia ter sido implementada por fatores que intuitivamente eliminam a responsabilidade, caso em que a presença da estrutura relevante não é em si mesma suficiente para a responsabilização (FISCHER; RAVIZZA, 1998, P. 196-201; LOCKE, 1975; SLOTE, 1980).

Fischer e Ravizza argumentam que

> se o entrosamento [entre os desejos de ordem superior e inferior] fosse produzido por [...] lavagem cerebral ou publicidade subliminar [...] não consideraríamos a agente como moralmente responsável por seu comportamento[...]

porque o mecanismo psicológico que produziu o comportamento não seria, "em um sentido intuitivo importante, da **própria agente**." (1998, p. 197, grifo do autor). Em resposta a tal preocupação, Fischer e Ravizza argumentam que a responsabilidade

tem um componente histórico-pessoal importante, o qual os autores tentam capturar com sua teoria de como as agentes podem "se responsabilizar" pelo mecanismo psicológico que produz o seu comportamento (1998, p. 207-239). Para críticas da teoria de Fischer e Ravizza sobre a responsabilidade, veja Levy (2011, p. 103-106) e Pereboom (2001, p. 120-22). Já para teorias mais diferentes sobre a responsabilidade, consulte Enoch (2012), Mason (2019, p. 179-207) e Wolf (2001). Para trabalhos sobre a significância geral de histórias pessoais para a responsabilidade, veja Christman (1991), Vargas (2006) e D. Zimmerman (2003).

Parte da motivação de Fischer e Ravizza para o desenvolvimento da teoria da responsabilização foi garantir que agentes que foram manipuladas de certas maneiras não acabem sendo responsabilizadas. Diversos exemplos e argumentos, incluindo o tipo de manipulação que preocuparam Fischer e Ravizza, desempenharam um importante papel na literatura recente sobre a responsabilidade. Um deles é o exemplo de Bete/Ana, criado por Alfred Melle (1995, 2006b), que enfatiza as dificuldades confrontadas pelas teorias da responsabilidade que evitam as condições histórico-pessoais. No exemplo, Ana adquiriu suas preferências e valores de modo normal, mas Bete é manipulada por um grupo de neurocientistas para que tenha preferências e valores que sejam idênticos aos de Ana. Depois da manipulação, Bete é capaz de refletir sobre seus novos valores e, quando o faz, endossa-os entusiasticamente. Mas apesar de podermos normalmente considerar tal endossamento como um sinal de autocontrole associado à responsabilidade, Mele sugere que Bete, diferentemente de Ana, exibe meramente "[um simulacro de] autocontrole", visto que os novos valores de Bete lhe foram impostos (1995, p. 155). E se certos tipos de histórias pessoais similarmente eliminam a habilidade de uma agente para governar seu comportamento de modo genuíno ou autêntico, então agentes com tais histórias não serão moralmente responsáveis. Para réplicas a Mele e *insights* gerais sobre os casos de manipulação, confira Arpaly (2003), King (2013), McKenna (2004) e Todd (2011). Para a discussão das questões sobre a identidade pessoal que são levantadas em casos de manipulação, veja os trabalhos de Khoury (2013), Matheson (2014) e Shoemaker (2012).

No caso de Bete, alguém poderia adotar uma postura radical. Tal postura poderia envolver a observação de que, apesar de Bete ter adquirido seus novos valores de um modo estranho, e de uma forma que envolveu erros morais cometidos por ela, **todas as pessoas** adquiriram seus valores de modos que não estão totalmente sob o controle delas. De fato, seguindo a linha de argumentação de Galen Strawson (1994), descrito em § 3.3.2, pode-se observar que **ninguém** tem

controle último sobre seus valores, e mesmo se agentes normais tivessem alguma capacidade de direcionar e alterar seus valores, os fatores disposicionais que governam, o modo como essa capacidade é usada, são ultimamente o resultado de fatores que vão além do controle da agente. Dessa forma, talvez não seja tão claro, como pode aparentar à primeira vista, que Bete se distingua de agentes normais em termos de seus poderes de autocontrole e de sua responsabilidade moral por seu comportamento. No entanto, esse raciocínio pode ser uma faca de dois gumes: em vez de mostrar que Bete é assimilada à classe de agentes normais, agentes responsáveis, o raciocínio pode mostrar que agentes normais são assimilados à classe de agentes não responsáveis, como Bete. O Argumento dos Quatro Casos de Derk Pereboom emprega um movimento nesses termos (1995, 2001, 2007, 2014).

O argumento de Pereboom apresenta o Professor Plum em quatro diferentes cenários. Em cada cenário, Plum mata uma mulher chamada White, satisfazendo as condições de responsabilidade-moral-envolvendo-merecimento, mais comumente propostas por compatibilistas (e descritas nas seções anteriores deste verbete): Plum mata White porque ele quer e, apesar desse desejo estar de acordo com o caráter de Plum, ele não é irresistível. Assim, Plum também endossa seu desejo de matar White com base em uma perspectiva volitiva de ordem superior. Finalmente, Plum é moralmente competente no geral e o processo de deliberação que o leva à decisão de matar White é apropriadamente sensível a razões.

No primeiro caso, Plum é "criado por neurocientistas, que [...] o manipulam diretamente por meio do uso de tecnologia à rádio." (Pereboom 2001, p. 112). Esses cientistas fazem com que o raciocínio de Plum tome certo caminho, sensível às razões, que culmina em sua conclusão de que as razões egoístas em favor de matar White superam as razões para não a matar. Pereboom acredita que, neste caso, Plum é claramente não responsável por matar White, visto que seu comportamento foi determinado pelas ações dos neurocientistas. No segundo e terceiro casos, Plum é causalmente determinado a empreender o mesmo processo de raciocínio, como no primeiro, mas, no segundo, Plum é meramente programado para fazê-lo pelos neurocientistas, ao invés de ter sido criado por eles. No Caso 3, por sua vez, o raciocínio de Plum é o resultado de sua influência sócio-histórica que determina seu caráter. No quarto caso, Plum é apenas um ser humano normal, em um universo causalmente determinado, e ele decide matar White do mesmo modo que nos casos anteriores.

Pereboom afirma que não há diferença relevante entre os Casos 1, 2 e 3 tal que nossos julgamentos sobre a responsabilidade de Plum deveriam ser diferentes nos três casos. Ademais, a razão de que Plum não é responsável parece ser que,

em cada caso, seu comportamento é causalmente determinado por forças para além de seu controle (PEREBOOM, 2001, p. 116). Mas, então, devemos concluir que Plum não é responsável no quarto caso, tendo em vista que o determinismo causal é característica desse caso. Plum é apenas um ser humano normal em um universo causalmente determinado, a conclusão que podemos tirar sobre ele deveria se estender a todas as outras pessoas normais em universos causalmente determinados. Para um importante argumento de manipulação relacionado, observe o "Argumento Zigoto" em Mele (1995, 2006b, 2008).

O argumento de Pereboom inspirou uma série de objeções, como, por exemplo, a de que poderia ser argumentado que, no Caso 1, a manipulação da qual Plum é sujeito elimina sua responsabilidade por alguma razão, além do fato de que a manipulação causalmente determina o seu comportamento, o que poderia parar a generalização do Caso 1 aos demais casos subsequentes (FISCHER, 2004; MELE, 2005; DEMETRIOU, 2010). Para uma resposta a esta linha de argumentação, *vide* SHABO, 2010). Por outro lado, poderia ser argumentado, sobre bases compatibilistas, que Plum é responsável, no que diz respeito ao Caso 4, e essa conclusão poderia ser estendida aos casos anteriores, visto que Plum preenche as mesmas condições favoráveis ao compatibilismo sobre a responsabilidade naqueles casos. (MCKENNA, 2008).

O Argumento dos Quatro Casos tenta mostrar que se o determinismo é verdadeiro, então não podemos ser a fonte de nossas ações do modo exigido pela responsabilidade moral. Trata-se, portanto, de um argumento em prol do incompatibilismo, não um argumento em prol do ceticismo sobre responsabilidade moral. Mas, em combinação com o argumento de Pereboom, no qual carecemos do tipo de livre-arbítrio requerido para a responsabilidade, mesmo se o determinismo for falso (2001, p. 38-88; 2014, p. 30-70), o Argumento dos Quatro Casos emergiu com uma parte importante de uma detalhada e influente perspectiva cética. Para outras teorias céticas, ver Caruso (2016), Smilansky (2000), Waller (2011) e, também, o verbete sobre **ceticismo sobre a responsabilidade moral**.

3.3.4. A Condição Epistêmica sobre Responsabilidade

A condição epistêmica tem sido objeto de bastante interesse contemporâneo sobre a responsabilidade (como oposta à condição de liberdade ou controle, que está no centro do debate sobre o livre-arbítrio). Neste contexto, foi desenvolvido

um argumento epistêmico a favor do ceticismo acerca da responsabilidade. Em certo aspecto estrutural, o argumento se assemelha ao argumento cético de Galen Strawson (§ 3.3.2).

Às vezes, as agentes agem sob ignorância das prováveis más consequências de suas ações, e às vezes suas ignorâncias as desculpam por agirem dessa forma. Mas em outros casos, a ignorância de uma agente pode não a desculpar. Como podemos distinguir os casos nos quais a ignorância nos desculpa dos casos em que não nos desculpa? Uma proposta é que a ignorância deixa de nos desculpar quando a agente poderia ser culpada precisamente por ser ignorante. E uma proposta de quando a ignorância é culpável é que ela resulta de um ato condenável de **negligência** em que uma agente culposamente debilita, ou não consegue melhorar, sua própria posição epistêmica (H. SMITH, 1983). Em tal caso, a ignorância da agente parece ser sua própria culpa, de modo que não se pode usá-la a fim de desculpar a agente.

Mas quando uma pessoa negligente é condenável? Diversas filósofas e diversos filósofos sugeriram que somos culpados por atos negligentes somente quando nos engajamos com eles sabendo o que estamos fazendo e sabendo que não deveríamos fazê-lo (LEVY, 2011; ROSEN, 2004; M. ZIMMERMAN, 1997). Em última instância, a sugestão é que a ignorância pela qual alguém é culpável, e que leva a transgressões não conscientes culpáveis, tem sua fonte no **conhecimento** do comportamento errado. Assim, se alguém faz algo errado não conscientemente, então esta pessoa será culpável somente se pudermos explicar a sua carência do conhecimento (sua "não consciência") por referência a outra coisa que ela, **conscientemente**, fez errado.

Considere um exemplo de Gideon Rosen (2004) no qual uma cirurgiã dá ordens para que sua paciente receba a transfusão do tipo de sangue errado, e suponha que a cirurgiã não esteja ciente de que está cometendo um equívoco. De acordo com Rosen, a cirurgiã será culpabilizada por ferir sua paciente apenas se ela for culpabilizada por ser ignorante sobre o tipo de sangue da paciente, quando ela solicitar a transfusão, e, por sua vez, será culpabilizada apenas se sua ignorância se originar de alguma instância na qual a cirurgiã, conscientemente, deixou de fazer algo que deveria ter feito para evitar essa ignorância. Não será suficiente, por exemplo, explicar a ignorância da cirurgiã por sua falha em checar o relatório clínico da paciente. Para fundamentar a culpa, essa omissão por parte da cirurgiã deve ser em si mesma culpável, o que requeriria que a cirurgiã soubesse que se trata de um erro. E se a cirurgiã não estivesse ciente de que estava cometendo uma

grave omissão (quando deixou de conferir o relatório clínico da paciente), então, **esta** falha de conhecimento deve ser explicada por algum ato ou omissão culpável, isto é, uma ação ou omissão consciente. Por fim, para Rosen,

> o único ***locus* possível de responsabilidade original** [por um ato não consciente posterior] **é um ato acrático** [...] um pecado consciente. (2004, p. 307, grifo do autor).

De modo semelhante, Michael Zimmerman argumenta que

> toda culpabilidade pode ser traçada de volta a uma culpabilidade que envolve a falta de ignorância, ou seja, que envolve uma crença por parte da agente de que ele ou ela está fazendo algo moralmente errado. (1997, p. 418).

O raciocínio acima pode ser aplicado não apenas aos casos nos quais uma pessoa é inconsciente das consequências de sua ação, mas também nos casos em que uma pessoa é inconsciente do estatuto moral de seu comportamento. Um proprietário de escravos, por exemplo, poderia pensar que escravizar é permissível, e, portanto, com base na teoria aqui considerada, ele será culpável somente se for culpável por sua ignorância sobre o estatuto moral da escravidão, o que exigirá, por exemplo, que ele ignore evidências sobre seu estatuto moral, mesmo sabendo que isso é algo que não se deve fazer (ROSEN, 2003, 2004).

Essas reflexões fazer surgir duas formas de ceticismo sobre a responsabilidade moral (e, particularmente, sobre a culpabilidade). Primeiramente, podemos endossar uma forma de ceticismo epistêmico, com base no fato de que raramente temos um *insight* sobre se um malfeitor é acrático ou se uma malfeitora é acrática, isto é, se era um malfeitor ou uma malfeitora consciente, em um ponto mais apropriado na etiologia de uma dada ação (ROSEN, 2004). Por outro lado, ou acrescentando ao ponto anterior, alguém pode endossar uma forma mais substancial de ceticismo, com base no fato de que grande parte dos malfeitores e malfeitoras típicos não exibe o tipo de erro consciente supostamente requerido para a responsabilização. Dito de outro modo, talvez muitos malfeitores e muitas malfeitoras não saibam que assim o sejam **e** sua ignorância sobre isso não é culpa deles ou delas, visto que a

culpa não surge de uma instância prévia apropriada de erro consciente. Neste caso, muitos malfeitores e muitas malfeitoras comuns podem deixar de serem responsáveis por seu comportamento. Para sugestões céticas nestes termos, confira M. Zimmerman (1997) e Levy (2011).

Há mais para a dimensão epistêmica da responsabilidade do que aquilo que está contido no argumento cético acima, mas o argumento revela muito do que é de interesse neste domínio. Uma razão disso é de que se baseia, proeminentemente, em uma estratégia de **rastreamento**. Essa estratégia é usada, por exemplo, em teorias que apresentam uma pessoa que não preenche, no momento da ação, as condições de controle ou conhecimento da responsabilização, mas que, não obstante, parece ser responsável por seu comportamento. Em tal caso, a responsabilidade da agente pode ser fundamentada no fato de que sua falha em preencher certas condições da responsabilidade é rastreável por meio de ações anteriores empreendidas pela agente quando preenchia tais condições. Por exemplo, uma pessoa pode estar tão intoxicada que carece de controle (ou consciência) de seu comportamento, e ainda assim, pode ser apropriadamente responsável por seu comportamento intoxicado, na medida em que ela livremente encontrou os meios para se intoxicar. A estratégia de rastreamento desempenha um papel importante em muitas teorias da responsabilidade (*vide* FISCHER; RAVIZZA, 1998, p. 49-51), mas também está sujeita a críticas significativas (*vide* VARGAS, 2005). Para uma réplica, confira Fischer e Tognazzini (2009) e, para mais sobre o rastreamento, veja Khoury (2012), King (2014), Shabo (2015) e Timpe (2011).

As várias estratégias para rejeitar o argumento cético acima também ilustram posturas que podem ser tomadas sobre a relevância do conhecimento para a responsabilização, e tipicamente envolvem rejeitar a afirmação de que o erro consciente é fundamental para a culpabilidade. A esse respeito, por exemplo, pode-se argumentar que, muitas vezes, é moralmente imprudente executar ações quando uma pessoa está meramente incerta sobre a precisão delas, e que essa imprudência é suficiente para culpabilidade (*vide* GUERRERO, 2007; NELKIN; RICKLESS, 2017B; ROBICHAUD, 2014). Outra estratégia seria argumentar que a culpabilidade pode ser fundamentada em casos de más ações moralmente ignorantes se for razoável esperar que o malfeitor ou a malfeitora tivesse evitado sua ignorância moral e, particularmente, se sua ignorância é ela própria causada pelos vícios epistêmicos e morais da agente (FITZPATRICK, 2008, 2017). Relacionado a isso, pode-se argumentar que aquela pessoa, que não tem consciência de que o que faz é errado, seja culpada se possuir capacidades relevantes para evitar sua falha

de consciência. Essa abordagem pode ser particularmente promissora em casos em que a falta de consciência moral de uma agente se origina de uma falha em **lembrar** seus deveres morais (CLARKE, 2014, 2017; SHER, 2006b, 2009; *vide* RUDY-HILLER, 2017). Finalmente, pode-se simplesmente afirmar que malfeitores ou malfeitoras, moralmente ignorantes, podem nutrir, e expressar através de seu comportamento, atitudes questionáveis ou qualidades da vontade que são suficientes para a culpabilidade (ARPALY, 2003; BJÖRNSSON, 2017; HARMAN, 2011; MASON, 2015; TALBERT, 2013). Essa abordagem pode ser muito promissora em casos em que um malfeitor ou uma malfeitora está consciente dos resultados materiais de sua conduta, mas inconsciente do erro que provoca esses efeitos.

Para saber mais sobre o assunto, consulte o verbete **condição epistêmica para responsabilidade moral**.

Referências

Os Volumes especiais da *Midwest Studies in Philosophy*, citados na Introdução, são Volume 30 (2006) e Volume 38 (2014), Minneapolis: University of Minnesota Press.

ADAMS, R. M. Involuntary Sins. **The Philosophical Review**, v. 94, n. 1, p. 3-31. DOI: 10.2307/21847131985.

ARISTOTELE. **Nicomachean Ethics**. Indianapolis: Hackett, 1999.

ARNESON, R. The Smart Theory of Moral Responsibility and Desert. *In:* OLSARETTI, S. (ed.). **Desert and Justice.** Oxford: Clarendon Press, 2003, p. 233-258.

ARPALY, N. **Unprincipled Virtue**: An Inquiry Into Moral Agency. Oxford: Oxford University Press, 2003. DOI: 10.1093/0195152042.001.0001.

AYER, A. J. **Freedom and Necessity, in his Philosophical Essays.** London: MacMillan, 1954, p. 271-284.

BALAGUER, M. **Free Will as an Open Scientific Problem**. Cambridge: MIT Press, 2010.

BENNETT, J. Accountability. *In:* VAN STRAATEN, Z. (ed.). **Philosophical Subjects***:* Essays Presented to P. F. Strawson. Oxford: Oxford University Press, 1980, p. 59-80.

BENSON, P. Freedom and Value. **The Journal of Philosophy**, v. 84, n. 9, p. 465-486, 1987. DOI: 10.2307/2027060.

BENSON, P. Culture and Responsibility: A Reply to Moody-Adams. **Journal of Social Philosophy**, v. 32, n. 4, p. 610-620, 2001. DOI: 10.1111/0047-

2786.00117.

BJÖRNSSON, G. Explaining Away Epistemic Skepticism about Culpability. *In:* SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. *v. 4.* New York: Oxford University Press, 2017, p. 141-162.

BOBZIEN, S. **Determinism and Freedom in Stoic Philosophy**. Oxford: Oxford University Press, 1998. DOI: 10.1093/0199247676.001.0001.

BRINK, D. O.; NELKIN, D. K. Fairness and the Architecture of Responsibility1. *In:* Shoemaker, D. (ed.). **Oxford Studies in Agency and Responsibility**. v. 1. New York: Oxford University Press, 2013, p. 284-314. DOI: 10.1093/acprof:oso/9780199694853.003.0013.

BROAD, C. D. **Determinism, Indeterminism, and Libertarianism:** An Inaugural Lecture. Cambridge: Cambridge University Press, 1934.

CARUSO, G. D. Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model (Presidential Address). **Southwest Philosophy Review**, v. 32, n. 1, p. 25-48, 2016. DOI: 10.5840/swphilreview20163214.

CHISHOLM, R. Human Freedom and the Self. *In:* WATSON, G. (ed.). **Free Will**. 2. ed. New York: Oxford University Press, 2003, p. 26–37.

CHRISTMAN, J. Autonomy and Personal History. **Canadian Journal of Philosophy**, v. 21, n. 1, p. 1-24, 1991. DOI: 10.1080/00455091.1991.10717234.

CLARKE, R. **Libertarian Accounts of Free Will.** New York: Oxford University Press, 2003. DOI: 10.1093/019515987X.001.0001.

CLARKE, R. On an Argument for the Impossibility of Moral Responsibility. **Midwest Studies in Philosophy**, v. 29, p. 13-24, 2005. DOI: 10.1111/j.1475-4975.2005.00103.x.

CLARKE, R. Dispositions, Abilities to Act, and Free Will: The New Dispositionalism. **Mind**, v.118, n. 470, p. 323-351, 2009. DOI: 10.1093/mind/fzp034.

CLARKE, R. **Omissions:** Agency, Metaphysics, and Responsibility. New York: Oxford University Press, 2014. DOI: 10.1093/acprof:oso/9780199347520.001.0001.

CLARKE, R. Blameworthiness and Unwitting Omissions. *In:* NELKIN, D. K.; RICKLESS, Samuel C. (ed). **The Ethics and Law of Omissions**. Oxford: Oxford University Press, 2017a, p. 63-83. DOI: 10.1093/oso/9780190683450.001.0001.

COATES, D. J.; TOGNAZZINI, N. A. The Contours of Blame. 2013a. *In:* COATES, D. J.; TOGNAZZINI, N. A. (ed.). **Blame:** Its Nature and Norms. New York: Oxford University Press, 2013b, p. 3–26. DOI: 10.1093/acprof:oso/9780199860821.003.0001.

COATES, D. J.; TOGNAZZINI, N. A. (ed.). **Blame:** Its Nature and Norms. New York: Oxford University Press, 2013b. DOI: 10.1093/acprof:oso/9780199860821.001.0001.

DARWALL, S. **The Second-Person Standpoint:** Morality, Respect, and Accountability. Cambridge: Harvard University Press, 2006.

DEMETRIOU, M. K. The Soft-Line Solution to Pereboom's Four-Case Argument. **Australasian Journal of Philosophy**, v. 88, n. 4, p. 595-617, 2010. DOI: 10.1080/00048400903382691.

DORIS, J. M.; MURPHY, D. From My Lai to Abu Ghraib: The Moral Psychology of Atrocity. **Midwest Studies in Philosophy**, v. 31, p. 25-55, 2007. DOI: 10.1111/j.1475-4975.2007.00149.x.

DWORKIN, G. Acting Freely. **Noûs**, v. 4, n. 4, p. 367-383, 1970. DOI: 10.2307/2214680.

EKSTROM, L. **Free Will***:* A Philosophical Study. Boulder CO: Westview Press, 2000.

ENOCH, D. Being Responsible, Taking Responsibility, and Penumbral Agency. *In:* HEUER, U.; LANG, G. (ed.). **Luck, Value, and Commitment:** Themes From the Ethics of Bernard Williams. Oxford: Oxford University Press, 95-132, 2012. DOI: 10.1093/acprof:oso/9780199599325.003.0005.

ENOCH, D.; MARMOR, A. The Case Against Moral Luck. **Law and Philosophy**, v. 26, n. 4, p. 405-436, 2007. DOI: 10.1007/s10982-006-9001-3.

ESHLEMAN, A. Worthy of Praise: Responsibility and Better-than-Minimally-Decent Agency. *In:* SHOEMAKER, D.; TOGNAZZINI, N. (ed.). **Oxford Studies in Agency and Responsibility.** v. 2. New York: Oxford University Press, 2014, p. 216-242. DOI: 10.1093/acprof:oso/9780198722120.001.0001.

FARA, M. Masked Abilities and Compatibilism. **Mind,** v. 117, n. 468, p. 843-865, 2008. DOI: 10.1093/mind/fzn078.

FINE, C.; KENNETT, J. Mental Impairment, Moral Understanding and Criminal Responsibility: Psychopathy and the Purposes of Punishment. **International Journal of Law and Psychiatry,** v. 27, n.5, p. 425-443, 2004. DOI: 10.1016/j.ijlp.2004.06.005.

FISCHER, J. M. **The Metaphysics of Free Will:** An Essay on Control. Oxford: Blackwell, 1994.

FISCHER, J. M. Frankfurt-Style Compatibilism. *In:* BUSS, S.; OVERTON, L. (ed.). **Contours of Agency:** Essays on Themes from Harry Frankfurt. Cambridge: MIT Press, 2002, p. 1-26.

FISCHER, J. M. Responsibility and Manipulation. **The Journal of Ethics,** *v.* 8, n. 2, p. 145-177, 2004. DOI: 10.1023/B:JOET.0000018773.97209.84.

FISCHER, J. M. The Cards That Are Dealt You. **The Journal of Ethics**, v.10, n. 1–2, p. 107-129, 2006. DOI: 10.1007/s10892-005-4594-6.

FISCHER, J. M. The Frankfurt Cases: The Moral of the Stories. **The Philosophical Review**, v. 119, n. 3, p. 315-336, 2010. DOI: 10.1215/00318108-2010-002.

FISCHER, J. M.; KANE, R.; PEREBOOM, D.; VARGAS, M. (ed.). **Four Views on Free Will**. Oxford: Blackwell, 2007.

FISCHER, J. M.; RAVIZZA, M. Introduction. 1993a. *In:* FISCHER, J. M.; RAVIZZA, M. (ed.). **Perspectives on Moral Responsibility**. NY: Cornell University Press, 1993b, p.1-41.

FISCHER, J. M.; RAVIZZA, M. (ed.). **Perspectives on Moral Responsibility**. NY: Cornell University Press, 1993b.

FISCHER, J. M.; RAVIZZA, M. **Responsibility and Control**: A Theory of Moral Responsibility. Cambridge: Cambridge University Press,1998. DOI: 10.1017/CBO9780511814594.

FISCHER, J. M.; TOGNAZZINI, N. A. The Truth about Tracing. **Noûs**, v.43, n. 3, p. 531-556, 2009. DOI: 10.1111/j.1468-0068.2009.00717.x.

FITZPATRICK, W. J. Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge. **Ethics**, v. 118, n. 4, p. 589–613, 2008. DOI: 10.1086/589532.

FITZPATRICK, W. J. Unwitting Wrongdoing, Reasonable Expectations, and Blameworthiness. *In:* ROBICHAUD, P.; Wieland, J. W. (ed.). **Responsibility:** The Epistemic Condition. Oxford: Oxford University Press, 2017, p. 29-46.

FRANKFURT, H. G. Alternate Possibilities and Moral Responsibility. **The Journal of Philosophy**, v. 66, n. 23, p. 829-839, 1969. DOI: 10.2307/2023833.

FRANKFURT, H. G. Freedom of the Will and the Concept of a Person. **The Journal of Philosophy**, v. 68, n. 1, p. 5-20, 1971. DOI: 10.2307/2024717.

FRANKFURT, H. G. Identification and Wholeheartedness. *In:* SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 1987, p. 27-45. DOI: 10.1017/CBO9780511625411.002.

FRANKFURT, H. G. Some Thoughts Concerning PAP. *In:* WIDERKER, D.; MCKENNA, Michael (ed.). **Moral Responsibility and Alternative Possibilities:** Essays on the Importance of Alternative Possibilities. Burlington: Ashgate, 2006, p. 339-445.

FRANKLIN, C. E. **A Minimal Libertarianism**: Free Will and the Promise of Reduction. New York: Oxford University Press, 2018. DOI: 10.1093/oso/

9780190682781.001.0001.

FRICKER, M. The Relativism of Blame and Williams's Relativism of Distance. **Aristotelian Society Supplementary**, v. 84, p. 151–177, 2010. DOI: 10.1111/j.1467-8349.2010.00190.x.

FRICKER, M. What's the Point of Blame? A Paradigm Based Explanation. **Noûs**, v. 50, n. 1, p. 165-183, 2016. DOI: 10.1111/nous.12067.

GINET, C. Might We Have No Choice? *In:* LEHRER, K. (ed.). **Freedom and Determinism**. New York: Random House, 1966, p. 87-104.

GINET, C. In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing. **Philosophical Perspectives**, v. 10, p. 403-417, 1996.

GLANNON, W. Psychopathy and Responsibility. **Journal of Applied Philosophy**, v. 14, n. 3, p. 263-275, 1997. DOI: 10.1111/1468-5930.00062.

GREENSPAN, P. S. Responsible Psychopaths. **Philosophical Psychology**, v. 16, n.3, p. 417-429, 2003. DOI: 10.1080/0951508032000121797.

GUERRERO, A. A. Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution. **Philosophical Studies**, v. 136, n. 1, p. 59-97, 2007. DOI: 10.1007/s11098-007-9143-7.

HAJI, I. **Moral Appraisability**. New York: Oxford University Press, 1998.

HANNA, N. Moral Luck Defended: Moral Luck Defended. **Noûs**, v. 48, n. 4, p. 683-698, 2014. DOI: 10.1111/j.1468-0068.2012.00869.x.

HARMAN, E. Does Moral Ignorance Exculpate? **Ratio**, v. 24, n.4, p. 443-468, 2011. DOI: 10.1111/j.1467-9329.2011.00511.x.

HARTMAN, R. J. **Defense of Moral Luck**: Why Luck Often Affects Praiseworthiness and Blameworthiness. New York: Routledge, 2017.

HIERONYMI, P. Articulating an Uncompromising Forgiveness. **Philosophy and Phenomenological Research**, v. 62, n. 3, p. 529-555, 2001. DOI: 10.1111/j.1933-1592.2001.tb00073.x.

HIERONYMI, P. The Force and Fairness of Blame. **Philosophical Perspectives**, v. 18, n. 1, p. 115-148, 2004. DOI: 10.1111/j.1520-8583.2004.00023.x.

HIERONYMI, P. Responsibility for Believing. **Synthese**, v. 161, n. 3, p. 357–373, 2008. DOI: 10.1007/s11229-006-9089-x.

HIERONYMI, P. Reflection and Responsibility: Reflection and Responsibility. **Philosophy & Public Affairs**, v. 42, n. 1, p. 3-41, 2014. DOI: 10.1111/papa.12024.

HOBART, R. E. Free Will as Involving Determination and Inconceivable without It. **Mind**, v. 43, n. 169, p. 1-27, 1934. DOI: 10.1093/mind/XLIII.169.1.

HOBBES, T. Of Liberty and Necessity. *In:* CHAPPELL, V. (ed.). *In*: **Hobbes and Bramhall on Liberty and Necessity**. Cambridge: Cambridge University Press, 1999, p. 15-42.

HONDERICH, T. **How Free Are You?:** The Determinism Problem. Oxford: Oxford University Press, 2002.

HUME, D. **An Enquiry Concerning Human Understanding**. Oxford: Oxford University Press, 1978.

HUNT, D. P. Moral Responsibility and Unavoidable Action. **Philosophical Studies**, v. 97, n. 2, p. 195-227, 2000. DOI: 10.1023/A:1018331202006.

JEFFERSON, A. Instrumentalism about Moral Responsibility Revisited. **The Philosophical Quarterly**, v. 69, n. 276, p. 555-573, 2019. DOI: 10.1093/pq/pqy062.

KANE, R. **The Significance of Free Will**. New York: Oxford University Press, 1996. DOI: 10.1093/0195126564.001.0001.

KANE, R. Libertarianism. *In:* FISCHER, J. M.; KANE, R.; PEREBOOM, D.; VARGAS, M. (ed.). **Four Views on Free Will**. Oxford: Blackwell, 2007, p. 5-43.

KELLY, E. I. What Is an Excuse? 2013. *In:* COATES, D. J.; TOGNAZZINI, N. A (ed.). **Blame:** Its Nature and Norms. New York: Oxford University Press, 2013b, p. 244-262. DOI: 10.1093/acprof:oso/9780199860821.003.0013.

KENNETT, J. Competence, Attributability, and Blame: Resolving the Responsibility of the Psychopath. *In:* COATES, J.; TOGNAZZINI, N. A. (eds.). **Oxford Studies in Agency and Responsibility.** Oxford: Oxford University Press, 2019. v. 5. p. 142-164. DOI: 10.1093/oso/9780198830238.003.0008.

KHOURY, A. C. Responsibility, Tracing, and Consequences. ***Canadian Journal of Philosophy*****,** v. 42, n. 3-4, p. 187-207, 2012. DOI: 10.1080/00455091.2012.10716774.

KHOURY, A. C. Synchronic and Diachronic Responsibility. **Philosophical Studies**, v. 165, n.3, p. 735-752 , 2013. DOI: 10.1007/s11098-012-9976-6.

KHOURY, A. C. The Objects of Moral Responsibility. **Philosophical Studies**, v. 175, n. 6, p. 1357-138, 2018. DOI: 10.1007/s11098-017-0914-5.

KING, M. The Problem with Manipulation. **Ethics**, v. 124, n.1, p. 65-83, 2013. DOI: 10.1086/671391.

KING, M. Traction without Tracing: A (Partial) Solution for Control-Based Accounts of Moral Responsibility. **European Journal of Philosophy**, v. 22, n. 3, p.

463-482, 2014. DOI: 10.1111/j.1468-0378.2011.00502.x.

LEHRER, Keith. Cans without Ifs. **Analysis**, v. 29, n. 1, p. 29-32, 1968. DOI: 10.1093/analys/29.1.29.

LEVY, N. Cultural Membership and Moral Responsibility. **The Monist**, v. 86, n. 2, p. 145-163, 2003. DOI: 10.5840/monist200386211.

LEVY, N. The Good, the Bad, and the Blameworthy. **Journal of Ethics and Social Philosophy**, v. 1, n. 2, p. 1-16, 2005. DOI: 10.26556/jesp.v1i2.6.

LEVY, N. The Responsibility of the Psychopath Revisited. **Philosophy, Psychiatry, & Psychology**, v. 14, n. 2, p. 129-138, 2007. DOI: 10.1353/ppp.0.0003.

LEVY, N. Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick. **Ethics**, v. 119, n. 4, p. 729-741, 2009. DOI: 10.1086/605018.

LEVY, N. **Hard Luck:** How Luck Undermines Free Will and Moral Responsibility. New York: Oxford University Press, 2011. DOI: 10.1093/acprof:oso/9780199601387.001.0001.

LEVY, N. Skepticism and Sanction: The Benefits of Rejecting Moral Responsibility. **Law and Philosophy**, v. 31, n. 5, p. 477-493, 2012. DOI: 10.1007/s10982-012-9128-3.

LEWIS, D. Are We Free to Break the Laws? **Theoria**, v. 47, n.3, p. 113-121, 1981. DOI: 10.1111/j.1755-2567.1981.tb00473.x.

LITTON, P. Psychopathy and Responsibility Theory. **Philosophy Compass**, v. 5, n. 8, p. 676-688, 2010. DOI: 10.1111/j.1747-9991.2010.00316.x.

LOCKE, D. Three Concepts of Free Action: I. **Proceedings of the Aristotelian Society**, v. 49, p. 95-112, 1975.

MACNAMARA, C. Blame, Communication, and Morally Responsible Agency. *In*: CLARKE, R.; MCKENNA, M.; SMITH, A. M. (ed.). **The Nature of Moral Responsibility**: New Essays. New York: Oxford University Press, 2015. p. 211-236. DOI: 10.1093/acprof:oso/9780199998074.003.0010.

MAIBOM, H. L. The Mad, the Bad, and the Psychopath. **Neuroethics**, v. 1, n. 3, p. 167-184, 2008. DOI: 10.1007/s12152-008-9013-9.

MASON, E. Moral Ignorance and Blameworthiness. **Philosophical Studies**, v. 172, n. 11, p. 3037-3057, 2015. DOI:10.1007/s11098-015-0456-7

MASON, E. Moral Incapacity and Moral Ignorance. *In:* PEELS, Rik. (ed.). **Perspectives on Ignorance from Moral and Social Philosophy**. New York: Routledge, 2017, p. 30-52.

MASON, E. **Ways to Be Blameworthy**: Rightness, Wrongness, and Responsibility. Oxford: Oxford University Press, 2019. DOI: 10.1093/oso/

9780198833604.001.0001.

MATHESON, B. Compatibilism and Personal Identity. **Philosophical Studies**, v. 170, n. 2, p. 317-334, 2014. DOI: 10.1007/s11098-013-0220-9.

MATHESON, B. In Defence of the Four-Case Argument. **Philosophical Studies**, v. 173, n. 7, p. 1963-1982, 2016. DOI: 10.1007/s11098-015-0587-x.

MCGEER, V. Building a Better Theory of Responsibility. **Philosophical Studies**, v. 172, n. 10, 2015, p. 2635-2649. DOI: 10.1007/s11098-015-0478-1.

MCKENNA, M. Responsibility and Globally Manipulated Agents. **Philosophical Topics,** v. 32, n. 1/2, p. 169-192, 2004. DOI: 10.5840/philtopics2004321/222.

MCKENNA, M. Reasons Reactivity and Incompatibilist Intuitions. **Philosophical Explorations**, v. 8, n. 2, p. 131-143, 2005. DOI: 10.1080/13869790500091508.

MCKENNA, M. A Hard-Line Reply to Perebooms Four-Case Manipulation Argument. **Philosophy and Phenomenological Research**, v. 77, n. 1, p. 142-159, 2008. DOI: 10.1111/j.1933-1592.2008.00179.x.

MCKENNA, M. **Conversation and Responsibility**. New York: Oxford University Press, 2012. DOI: 10.1093/acprof:oso/9780199740031.001.0001.

MCKENNA, M. Reasons-Responsiveness, Agents, and Mechanisms. SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. New York: Oxford University Press, 2013. v. 1. p. 151-183. DOI: 10.1093/acprof:oso/9780199694853.003.0007.

MELE, A. R. **Autonomous Agents**: From Self-Control to Autonomy. New York: Oxford University Press, 1995. DOI:10.1093/0195150430.001.0001

MELE, A. R. A Critique of Pereboom's 'Four-Case Argument' for Incompatibilism. **Analysis**, v. 65, n. 1, p. 75-80, 2005. DOI: 10.1093/analys/65.1.75.

MELE, A. R. Fischer and Ravizza on Moral Responsibility. **The Journal of Ethics**, v. 10, n. 3, p. 283-294, 2006a. DOI: 10.1007/s10892-005-5780-2.

MELE, A. R. **Free Will and Luck.** New York: Oxford University Press, 2006b. DOI: 10.1093/0195305043.001.0001.

MELE, A. R. Manipulation, Compatibilism, and Moral Responsibility. **The Journal of Ethics**, v. 12, n. 3-4, p. 263-286, 2008. DOI: 10.1007/s10892-008-9035-x.

MELE, A. R.; ROBB, D. Rescuing Frankfurt-Style Cases. **The Philosophical Review**, v. 107, n. 1, p. 97-112, 1998. DOI: 10.2307/2998316.

MILAM, Per-Erik. Reactive Attitudes and Personal Relationships. **Canadian Journal of Philosophy**, v. 46, n. 1, p. 102-122, 2016. DOI: 10.1080/00455091.2016.1146032.

MOODY-ADAMS, M. On the Old Saw that Character is Destiny. *In:* FLANAGAN, O.; RORTY, A. O. (ed.). **Identity, Character, and Morality:** Essays in Moral Psychology. Cambridge: MIT Press, 1990, p.111-32.

MOORE, G. E. **Ethics**. Oxford: Oxford University Press, 1912.

NAGEL, T. Moral Luck. *In*: **Mortal Questions**. Cambridge: Cambridge University Press, 1979. p. 24-38. DOI: 10.1093/aristoteliansupp/50.1.115.

NEELY, W. Freedom and Desire. **The Philosophical Review**, v. 83, n. 1, p. 32-54, 1974. DOI: 10.2307/2183872.

NELKIN, D. K. Responsibility and Rational Abilities: Defending an Asymmetrical View. **Pacific Philosophical Quarterly**, v. 89, n. 4, p. 497–515, 2008. DOI: 10.1111/j.1468-0114.2008.00333.x.

NELKIN, D. K. **Making Sense of Free Will and Responsibility**. New York: Oxford University Press, 2011.

NELKIN, D. K. Psychopaths, Incorrigible Racists, and the Faces of Responsibility. **Ethics**, v. 125, n. 2, p. 357-390, 2015. DOI:10.1086/678372

NELKIN, D. K.; RICKLESS, S. C. (ed.). **The Ethics and Law of Omissions**. Oxford: Oxford University Press, 2017a. DOI: 10.1093/oso/9780190683450.001.0001.

NELKIN, D. K. Moral Responsibility for Unwitting Omissions: A New Tracing View. In: NELKIN, D. K.; RICKLESS, S. C. (ed.). **The Ethics and Law of Omissions**. Oxford: Oxford University Press, 2017a. p.106-130.

NICHOLS, S. How Psychopaths Threaten Moral Rationalism: Is It Irrational To Be Amoral?. **The Monist**, v. 85, n. 2, p. 285-303, 2002. DOI: 10.5840/monist200285210.

O'CONNOR, T. **Persons and Causes**: The Metaphysics of Free Will. New York: Oxford University Press, 2000. DOI: 10.1093/019515374X.001.0001.

PEELS, R. A Modal Solution to the Problem of Moral Luck. **American Philosophical Quarterly**, v. 52, n. 1, p. 73-87, 2015.

PEREBOOM, D. Determinism al Dente. **Noûs**, v. 29, n. 1, p. 21-45,1995. DOI: 10.2307/2215725.

PEREBOOM, D. Alternative Possibilities and Causal Histories. **Philosophical Perspectives**, v. 14, p. 119-137, 2000.

PEREBOOM, D. **Living Without Free Will**. Cambridge: Cambridge University Press, 2001. DOI: 10.1017/CBO9780511498824.

PEREBOOM, D. Hard Incompatibilism. *In:* FISCHER, J. M.; KANE, R.; PEREBOOM, D.; VARGAS, M. (ed.). **Four Views on Free Will**. Oxford: Blackwell, 2007, p. 85-125.

PEREBOOM, D. **Free Will, Agency, and Meaning in Life.** New York: Oxford University Press, 2014. DOI: 10.1093/acprof:oso/9780199685516.001.0001.

PEREBOOM, D. Responsibility, Regret, and Protest. In: SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. New York: Oxford University Press, 2017a. v. 4, p. 121-140.

ROBICHAUD, P. On Culpable Ignorance and Akrasia. **Ethics**, v. 125, n. 1, p. 137-151, 2014. DOI: 10.1086/677139.

ROSEN, G. Culpability and Ignorance. **Proceedings of the Aristotelian Society**, v. 103, n. 1, p. 61-84, 2003. DOI: 10.1111/j.0066-7372.2003.00064.x.

ROSEN, G. Skepticism about Moral Responsibility. **Philosophical Perspectives**, v. 18, p. 295-313, 2004.

RUDY-HILLER, F. A Capacitarian Account of Culpable Ignorance. **Pacific Philosophical Quarterly**, v. 98, supl. 1, p. 398-426, 2017. DOI: 10.1111/papq.12190.

RUSSELL, P. Strawson's Way of Naturalizing Responsibility. **Ethics**, v. 102, n. (2), p. 287–302, 1992. DOI: 10.1086/293397.

RUSSELL, P. **Freedom and Moral Sentiment:** Hume's Way of Naturalizing Responsibility. Oxford: Oxford University Press, 1995. DOI: 10.1093/0195152905.001.0001.

RUSSELL, P. Responsibility and the Condition of Moral Sense. **Philosophical Topics**, v. 32, n. 1/2, p. 287-305, 2004. DOI: 10.5840/philtopics2004321/24.

RUSSELL, P. Free Will Pessimism. *In:* SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. New York: Oxford University Press, 2017a. v. 4, p. 93-120.

SALLES, R. **The Stoics on Determinism and Compatibilism**. Burlington: Ashgate Publishing, 2005.

SARTORIO, C. **Causation and Free Will**. New York: Oxford University Press, 2016. DOI: 10.1093/acprof:oso/9780198746799.001.0001.

SCANLON, T. M. **What We Owe to Each Other**. Cambridge: Harvard University Press, 1998.

SCANLON, T. **Moral Dimensions**: Permissibility, Meaning, and Blame. Cambridge: Harvard University Press, 2008.

SCHLICK, M. When is a Man Responsible? *In*: BEROFSKY, B. (ed.). **Free Will and Determinism**. New York: Harper & Row, 1966, p. 54-63.

SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 1987. DOI: 10.1017/CBO9780511625411.

SCHRAMME, T. (ed.). **Being Amoral**: Psychopathy and Moral Incapacity. Cambridge: MIT Press, 2014.

SHABO, S. Uncompromising Source Incompatibilism. **Philosophy and Phenomenological Research**, v. 80, n. 2, p. 349-383, 2010. DOI: 10.1111/j.1933-1592.2010.00328.x.

SHABO, S. More Trouble with Tracing. **Erkenntnis,** v. 80, n. 5, p. 987-1011, 2015. DOI: 10.1007/s10670-014-9693-y.

SHER, G. **Praise of Blame**. New York: Oxford University Press, 2006a. DOI: 10.1093/0195187423.001.0001.

SHER, G. Out of Control. **Ethics**, v. 116, n. 2, p. 285-301, 2006b. DOI: 10.1086/498464.

SHER, G. **Who Knew? Responsibility Without Awareness**. New York: Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780195389197.001.0001.

SHOEMAKER, D. Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. **Ethics**, v. 121, n. 3, p. 602-632, 2011. DOI: 10.1086/659003.

SHOEMAKER, D. Responsibility Without Identity. **The Harvard Review of Philosophy**, v. 18, p. 109-132, 2012. DOI: 10.5840/harvardreview20121816.

SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. New York: Oxford University Press, 2013. v. 1. DOI: 10.1093/acprof:oso/9780199694853.001.0001.

SHOEMAKER, D. (ed.). Psychopathy, Responsibility, and the Moral/Conventional Distinction. *In:* SCHRAMME, Thomas (ed.). **Being Amoral**: Psychopathy and Moral Incapacity. Cambridge: MIT Press, 2014, p. 247-274.

SHOEMAKER, D. **Responsibility from the Margins**. New York: Oxford University Press, 2015a. DOI: 10.1093/acprof:oso/9780198715672.001.0001.

SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. New York: Oxford University Press, 2015b. v. 3. DOI: 10.1093/acprof:oso/9780198744832.001.0001.

SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. New York: Oxford University Press, 2017a. v. 4. DOI: 10.1093/oso/9780198805601.001.0001.

SHOEMAKER, D. Response-Dependent Responsibility: or A Funny Thing Happened on the Way to Blame. **Philosophical Review**, v. 126, n. 4, p. 481-527, 2017b. DOI: 10.1215/00318108-4173422.

SHOEMAKER, D.; TOGNAZZINI, N. (ed.). **Oxford Studies in Agency and Responsibility.** v. 2. New York: Oxford University Press, 2014. DOI: 10.1093/

acprof:oso/9780198722120.001.0001.

SLOTE, Michael A. Understanding Free Will. **The Journal of Philosophy**, v. 77, n. 3, p. 136-151, 1980. DOI: 10.2307/2025666.

SMART, J. J. C. Free-Will, Praise and Blame. **Mind**, v. 70, n. 279, p. 291-306, 1961. DOI: 10.1093/mind/LXX.279.291.

SMART, J. J. C. An Outline of a Utilitarian System of Ethics. *In:* SMART, J. J. C.; WILLIAMS, Bernard. **Utilitarianism:** For and Against. Cambridge: Cambridge University Press, 1973, p. 3-74.

SMILANSKY, Saul. **Free Will and Illusion**. New York: Oxford University Press, 2000.

SMITH, A. M. Responsibility for Attitudes: Activity and Passivity in Mental Life. **Ethics**, v. 115, n. 2, p. 236-271, 2005. DOI: 10.1086/426957.

SMITH, A. M. Control, Responsibility, and Moral Assessment. **Philosophical Studies**, v. 138, n. 3, p. 367-392, 2008. DOI: 10.1007/s11098-006-9048-x.

SMITH, A. M. Moral Blame and Moral Protest. 2013. *In:* COATES, D. J.; TOGNAZZINI, N. A. (ed.). **Blame:** Its Nature and Norms. New York: Oxford University Press, 2013b, p. 27-48. DOI: 10.1093/acprof:oso/9780199860821.003.0002.

SMITH, A. M. Responsibility as Answerability, **Inquiry**, v. 58, n. 2, p. 99-126, 2015. DOI: 10.1080/0020174X.2015.986851.

SMITH, H. Culpable Ignorance. **The Philosophical Review**, v. 92, n. 4, p. 543-5711983. DOI: 10.2307/2184880.

SMITH, M. Rational Capacities or How to Distinguish Recklessness, Weakness, and Compulsion. *In:* STROUD, Sarah; TAPPOLET, C. (ed.). **Weakness of Will and Practical Irrationality**. Oxford: Oxford University Press, 2003, p. 17-38. DOI: 10.1093/0199257361.003.0002.

SRIPADA, C. Self-Expression: A Deep Self Theory of Moral Responsibility. **Philosophical Studies**, v. 173, n. 5, p. 1203-1232, 2016. DOI: 10.1007/s11098-015-0527-9.

SRIPADA, C. Frankfurt's Unwilling and Willing Addicts. **Mind**, v. 126, n. 503, p. 781-815, 2017.DOI: 10.1093/mind/fzw013.

STRAWSON, G. **Freedom and Belief**. Oxford: Oxford University Press, 1986. DOI: 10.1093/acprof:oso/9780199247493.001.0001.

STRAWSON, G. The Impossibility of Moral Responsibility. **Philosophical Studies**, v. 75, n. 1-2, p. 5–24, 1994. DOI: 10.1007/BF00989879.

STRAWSON, P. F. Freedom and Resentment. 1993. *In:* FISCHER, J. M.; RAVIZZA, M. (ed.). **Perspectives on Moral Responsibility**. NY: Cornell University Press, 1993b. p. 45-66.

TALBERT, M. Moral Competence, Moral Blame, and Protest. **The Journal of Ethics**, v. 16, n. 1, p. 89-109, 2012. DOI: 10.1007/s10892-011-9112-4.

TALBERT, M. Unwitting Wrongdoers and the Role of Moral Disagreement in Blame. *In:* SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. New York: Oxford University Press, 2013. v. 1, p. 225-245. DOI: 10.1093/acprof:oso/9780199694853.003.0010.

TALBERT, M. The Significance of Psychopathic Wrongdoing. *In:* SCHRAMME, T. (ed.). **Being Amoral**: Psychopathy and Moral Incapacity. Cambridge: MIT Press, 2014. p. 275-300.

TAYLOR, C. Responsibility for Self. *In*: RORTY, A. O. (ed.). **The Identities of Persons**. Berkeley: University of California Press, 1976. p. 281-99.

TIMPE, K. Tracing and the Epistemic Condition on Moral Responsibility. **The Modern Schoolman**, v. 88, n. 1-2, p. 5-28. 2011. DOI: 10.5840/schoolman2011881/22.

TODD, P. A New Approach to Manipulation Arguments. **Philosophical Studies**, v. 152, n. 1, p. 127-133, 2011. DOI: 10.1007/s11098-009-9465-8.

TODD, P. Strawson, Moral Responsibility, and the 'Order of Explanation': An Intervention. **Ethics**, v. 127, n.1, p. 208-240, 2016. DOI: 10.1086/687336.

TODD, P.; TOGNAZZINI, N. A. A Problem for Guidance Control. **The Philosophical Quarterly**, v. 58, n. 233, p. 685-692, 2008. DOI: 10.1111/j.1467-9213.2008.576.x.

VAN INWAGEN, P. **An Essay on Free Will**. New York: Oxford University Press, 1983.

VARGAS, M. The Trouble with Tracing. **Midwest Studies in Philosophy**, v. 29, p. 269-291, 2005. DOI: 10.1111/j.1475-4975.2005.00117.x.

VARGAS, M. On the Importance of History for Responsible Agency. **Philosophical Studies**, v. 127, n. 3, p. 351-382, 2006. DOI: 10.1007/s11098-004-7819-9.

VARGAS, M. **Building Better Beings**: A Theory of Moral Responsibility. New York: Oxford University Press, 2013. DOI: 10.1093/acprof:oso/9780199697540.001.0001.

VARGAS, M.; NICHOLS, S. Psychopaths and Moral Knowledge. **Philosophy, Psychiatry, & Psychology**, v.14, n.2, p. 157-162, 2007. DOI: 10.1353/ppp.0.0000.

VIHVELIN, K. Free Will Demystified: A Dispositional Account. **Philosophical Topics**, v. 32, n. 1/2, p. 427-450, 2004. DOI: 10.5840/philtopics2004321/211.

WALLACE, R. J. **Responsibility and the Moral Sentiments**. Cambridge: Harvard University Press, 1996.

WALLACE, R. J. Dispassionate Opprobrium: On Blame and the Reactive Sentiments. *In:* WALLACE, R.; RAHUL, K. J.; FREEMAN, S. **Reasons and Recognition:**

Essays on the Philosophy of T.M. New York: Oxford University Press, 2011. p. 348-372. DOI: 10.1093/acprof:oso/9780199753673.001.0001.

WALLER, B. **Against Moral Responsibilit**. Cambridge: MIT Press, 2011.

WATSON, G. Free Agency. **The Journal of Philosophy**, v. 72, n. 8, p. 205-220, 1975. DOI: 10.2307/2024703.

WATSON, G. Soft Libertarianism and Hard Compatibilism. **The Journal of Ethics**, v. 3, n. 4, p. 353-368, 1999. DOI: 10.1023/A:1009819618482.

WATSON, G. Reasons and Responsibility. **Ethics**, v. 111, n. 2, p. 374-394, 2001. DOI: 10.1086/233477.

WATSON, G. Contractualism and the Boundaries of Morality: Remarks on Scanlon's What We Owe to Each Other. **Social Theory and Practice**, v. 28, n. 2, p. 221-241, 2002. DOI: 10.5840/soctheorpract20022829.

WATSON, G. Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. *In:* SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 1987, p. 219-259. DOI: 10.1017/CBO9780511625411.011.

WATSON, G. Two Faces of Responsibility. **Philosophical Topics**, v. 24, n. 2, p. 227-248, 1996. DOI: 10.5840/philtopics199624222.

WATSON, G. **Agency and Answerability**: Selected Essays. New York: Oxford University Press, 2004. DOI: 10.1093/acprof:oso/9780199272273.001.0001.

WATSON, G. The Trouble with Psychopaths. *In*: WALLACE, R.; RAHUL, K. J.; FREEMAN, S. **Reasons and Recognition:** Essays on the Philosophy of T.M. New York: Oxford University Press, 2011, p. 307-31. DOI: 10.1093/acprof:oso/9780199753673.001.0001.

WIDERKER, D. Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. **The Philosophical Review**, v. 104, n. 2, p. 247-261, 1995. DOI: 10.2307/2185979.

WIGGINS, D. Towards a Reasonable Libertarianism. *In:* HONDERICH, T. (ed.). **Essays on Freedom of Action**. London: Routledge & Kegan Paul, 1973, p. 31-62.

WILLIAMS, B. Moral Luck. *In:* **Proceedings of the Aristotelian Society Supplementary**, v. 50, p. 115-135, 1976. DOI: 10.1093/aristoteliansupp/50.1.115.

WOLF, S. Asymmetrical Freedom. **The Journal of Philosophy**, v. 77, n. 3, p. 151-166, 1980. DOI: 10.2307/2025667.

WOLF, S. Sanity and the Metaphysics of Responsibility. *In:* SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral

Psychology. Cambridge: Cambridge University Press, 1987, p. 46-62. DOI: 10.1017/CBO9780511625411.003.

WOLF, S. **Freedom Within Reason**. New York: Oxford University Press, 1990.

WOLF, S. The Moral of Moral Luck. **Philosophic Exchange**, v. 31, p. 4-19, 2001.

WOLF, S. Blame, Italian Style. *In:* WALLACE, R.; RAHUL, K. J.; FREEMAN, S. **Reasons and Recognition:** Essays on the Philosophy of T.M. New York: Oxford University Press, 2011, p. 332-347.

ZIMMERMAN, D. That Was Then, This Is Now: Personal History vs. Psychological Structure in Compatibilist Theories of Autonomous Agency. **Noûs**, v. 37, n. 4, p. 638-671, 2003. DOI: 10.1046/j.1468-0068.2003.00454.x.

ZIMMERMAN, M. J. **An Essay on Moral Responsibility**. Totowa: Rowman and Littlefield, 1988.

ZIMMERMAN, M. J. Moral Responsibility and Ignorance. **Ethics**, v. 107, n. 3, p. 410-426, 1997. DOI: 10.1086/233742.

ZIMMERMAN, M. J. Taking Luck Seriously. **The Journal of Philosophy**, v. 99, n.11, p. 553-576, 2002. DOI: 10.2307/3655750.

ZIMMERMAN, M. J. Moral Luck Reexamined. *In:* SHOEMAKER, D. **Responsibility from the Margins**. New York: Oxford University Press, 2015, p.136-159. DOI: 10.1093/acprof:oso/9780198715672.001.0001.

(III) Compatibilismo*

Autores: Michael McKenna & D. Justin Coates
Tradutores: Alex Bispo & Pedro Merlussi
Revisores Científicos: Karenina do Nascimento Rodrigues, Kherian Gracher & Rhamon de Oliveira Nunes

O Compatibilismo oferece uma solução ao **problema do livre-arbítrio**, que diz respeito ao debate sobre a incompatibilidade entre o livre-arbítrio e o determinismo. O **compatibilismo** é a tese de que o livre-arbítrio é compatível com o determinismo. Em geral, como o livre-arbítrio é considerado uma condição necessária para a responsabilidade moral, o compatibilismo, às vezes, é formulado como uma tese sobre a compatibilidade entre a responsabilidade moral e o determinismo.

---

* MCKENNA, M.; JUSTIN, D. Compatibilism. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. No Prelo. Stanford, CA: The Metaphysics Research Lab, 2021. Disponível em: https://plato.stanford.edu/archives/fall2021/entries/compatibilism/. Acesso em: 01 ago. 2021.

The following is the translation of the entry on Compatibilism, 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/fall2021/entries/compatibilism/>. 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/archives/win2019/entries/compatibilism/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry.

1. Livre-Arbítrio e o Problema do Determinismo Causal

O compatibilismo emerge como resposta a um problema posto pelo determinismo causal. Mas qual é o problema? Suponha que tudo que ocorre seja o resultado inevitável das leis da natureza e do estado do mundo no passado distante, como a tese do determinismo causal nos diz. Se este for o caso, então, tudo que agentes humanos fazem seguem as leis da natureza e a maneira como o mundo era no passado. Porém, se o que fazemos é simplesmente uma consequência das leis da natureza e do estado do mundo no passado, então, não poderíamos ter feito nada diferente do que realmente fizemos. Não seríamos, também, em qualquer sentido relevante, a origem causal última das nossas ações, uma vez que elas têm suas origens causais nas leis da natureza e no estado do mundo há muito tempo. Portanto, o determinismo parece impedir que agentes humanos tenham a liberdade para agir de outro modo, e parece também impedir que os agentes sejam autores de suas próprias ações. Se alguma dessas teses for verdadeira, então, é de se duvidar que agentes humanos sejam livres ou responsáveis por suas ações em qualquer sentido relevante.

Essas linhas de argumentação, que foram estabelecidas nas obras de Ginet (1996), van Inwagen (1975, 1983), Wisdom (1934), Mele (1995) e Pereboom (1995, 2001), dentre outros estudos, apresentam um problema real para aqueles ou aquelas que tendem a pensar que somos livres e responsáveis por nossas escolhas e ações, e que o mundo natural pode operar como um sistema determinístico (ou, se não completamente determinístico, um sistema no qual um indeterminismo é meramente um ruído estocástico que é causalmente irrelevante no nível da agência humana). Como responder a tais argumentos? Por um lado, as(os) **incompatibilistas** aceitam esses argumentos (ou ao menos uma versão de pelo menos um desses argumentos) e insistem que nossa autoconcepção como agentes livres e responsáveis estaria seriamente equivocada se o determinismo causal for verdadeiro. Algumas e alguns incompatibilistas alegam a favor dessas conclusões de maneira indireta – primeiro, argumentando que o determinismo impede liberdade ou controle e, depois, então, justificando que tal liberdade é necessária para responsabilidade moral. Por sua vez, outros incompatibilistas ainda argumentam diretamente que o determinismo causal exclui a responsabilidade moral.

Por outro lado, as(os) compatibilistas afirmam que essas preocupações não atingem o resultado desejado. Algumas e alguns compatibilistas defendem isso porque pensam que a verdade do determinismo causal não eliminaria nossa liberdade

para agir de outro modo (*vide* BEROFSKY, 1987; CAMPBELL, 1997; VIHVELIN, 2013; etc.). Como resultado, essas e esses compatibilistas defendem que a verdade do determinismo causal não representa uma ameaça ao nosso *status* de agentes moralmente responsáveis (note aqui a premissa que não foi expressa: a liberdade de agir de outro modo é suficiente para o tipo de controle que uma agente deve possuir para ser moralmente responsável por suas ações). Outras e outros compatibilista, por sua vez, demonstram menos preocupação em refutar a conclusão de que a liberdade para agir de outro modo seja incompatível com o determinismo. Esses rejeitam a ideia de que tal liberdade é necessária para formas relevantes de livre-arbítrio (FRANKFURT, 1969, 1971; WATSON, 1975; DENNETT, 1984), as chamadas "variedades de livre-arbítrio que valem a pena" (DENNETT, 1984). E ainda mais notavelmente, algumas e alguns compatibilistas simplesmente negam que esse tipo de liberdade seja de alguma forma relacionada à agência moralmente responsável (*vide* FISCHER, 1994; FISCHER; RAVIZZA, 1998; SCANLON, 1998; WALLACE, 1994; SARTORIO, 2016). O que vemos aqui não é uma frente unificada perante o desafio incompatibilista, pois, sendo cuidadosos ao identificar os elementos precisos dos argumentos incompatibilistas que consideram errados, eles constroem suas respectivas teorias sobre a liberdade e a responsabilidade com isso em mente.

Para ajudar a situar melhor as teorias compatibilistas, consideraremos com mais detalhes os argumentos que as(os) incompatibilistas desenvolveram em favor de suas próprias teorias, uma vez que são esses argumentos que moldaram as teorias compatibilistas. Como mencionamos acima, a verdade do determinismo causal, aparentemente, representa um problema para a liberdade e a responsabilidade em pelo menos duas maneiras. Em primeiro lugar, ela pode implicar que ninguém tem a **liberdade de agir de outro modo**, que é um tipo de poder ou controle sobre as ações de alguém que muitas pessoas consideraram como necessária para a responsabilidade moral. Em segundo lugar, ela pode implicar que ninguém é a origem causal última de suas ações. Uma condição que parece plausível para a agência moralmente responsável é que as nossas ações sejam originadas pelo livre-arbítrio, não em forças externas a nós. Separadamente, analisaremos cada uma dessas possíveis ameaças ao livre-arbítrio.

1.1. Determinismo e Possibilidades Alternativas

Uma maneira natural de pensarmos o controle de uma agente sobre a sua conduta em certo momento no tempo é em termos de sua capacidade de selecionar ou escolher entre cursos de ação alternativos. Essa imagem de controle surge dos aspectos comuns de nossas perspectivas como deliberadores práticos ou deliberadoras práticas que decidem sobre cursos de ação. Se uma pessoa está escolhendo entre votar em Clinton em oposição a Trump, é plausível assumir que a sua liberdade em relação ao seu voto consiste, ao menos parcialmente, em sua capacidade de escolher entre essas duas alternativas. De acordo com essa versão, agir com livre-arbítrio requer **possibilidades alternativas**. Uma maneira natural para tratar dessa versão do livre-arbítrio seria em termos do futuro de uma agente como um jardim de caminhos bifurcados, que se ramificam a partir de um único passado. O lugar de ação livre surge quando o presente, a partir de um passado único da agente, oferece mais de um caminho em direção ao futuro. Portanto, nesse modelo da agência humana, quando uma pessoa age de acordo com seu livre-arbítrio, **ela poderia ter agido de outro modo**.

Dessa concepção de livre-arbítrio imediatamente se segue a ideia de que o determinismo pode ser uma ameaça, pois o determinismo, entendido estritamente da maneira como foi caracterizado anteriormente, nos diz que, em qualquer tempo, há somente um futuro possível, em virtude dos fatos sobre o passado e as leis da natureza. Porém, de acordo com a concepção de agência humana em questão, uma agente livre poderia ter agido de maneira diferente da que agiu e, portanto, mais de um futuro seria possível.

Vejamos um argumento incompatibilista que demonstra as considerações discutidas acima:

A Uma agente qualquer, **x**, executa uma ação, **a**, de acordo com o seu livre-arbítrio se, e somente se, **x** tem controle sobre **a**.

B **x** tem controle sobre **a** somente se **x** tem a capacidade de escolher executar **a** entre cursos de ação alternativos.

C Se **x** tem a capacidade de escolher executar **a** entre cursos de ação alternativos, então existem cursos de ação alternativos para executar **a** disponíveis a **x** (ou seja, **x** poderia ter agido de modo diferente de **a**).

D Se o determinismo for verdadeiro, então somente um futuro é possível, mantendo fixos o passado e as leis da natureza.

E Se somente um futuro é possível, mantendo fixos o passado e as leis da natureza, então não existem cursos de ação alternativos para qualquer ação disponível para *x* (isto é, **x** não poderia ter agido de modo diferente do que ele atualmente agiu).
F Portanto, se o determinismo for verdadeiro, não é o caso que uma agente qualquer, **x**, execute uma ação qualquer, **a**, de acordo com o seu livre-arbítrio.

Para facilitar a referência e a discussão ao longo deste verbete, vamos simplificar o argumento acima da seguinte maneira:

1 Se alguma pessoa age de acordo com o seu livre-arbítrio, então ela poderia ter agido de outro modo (A-C).
2 Se o determinismo for verdadeiro, então ninguém pode agir de maneira diferente da que atualmente agiu (D-E).
3 Portanto, se o determinismo for verdadeiro, ninguém age de acordo com o seu livre-arbítrio (F).

Chamaremos esse argumento simplificado de **Argumento a favor do Incompatibilismo Clássico**. De acordo com o argumento, se o determinismo for verdadeiro, então ninguém tem acesso a alternativas do tipo que podemos, plausivelmente, considerar requeridas para se ter livre-arbítrio.

1.2. Determinismo e Origem

Há uma segunda concepção do tipo de controle que pode ser necessária para a agência moralmente responsável. Essa concepção começa com a observação de que o controle de uma agente parece consistir em sua ocupação de uma posição crucial na produção de suas ações. Pense em termos da diferença transparente entre aqueles eventos que **são produtos da agência de alguém** e aqueles que são meramente acontecimentos corporais. Por exemplo, considere a escolha de se pegar uma xícara de café em oposição a eventos como o batimento cardiáco ou a circulação sanguínea de alguém. No primeiro caso, a pessoa é a origem e a produtora do acontecimento. Nesse modelo de agência humana, controle é entendido em termos da agente ser a origem de suas ações. Focando nessa concepção de agência, como o determinismo pode representar uma ameaça ao livre-arbítrio? Se o determinismo for verdadeiro, então o que aconteceu no passado distante, quando

combinado com as leis da natureza, é causalmente suficiente para a produção de toda a ação humana. Mas se esse é o caso, então, embora possa ser verdade que uma agente esteja crucialmente envolvida na produção de sua ação, essa ação, na verdade, tem seu princípio em antecedentes causais que se **originam** fora da agente. Portanto, a agente não é a **origem última** de suas ações.

O que se entende aqui por origem última, e não apenas origem? Uma agente é a origem última de sua ação somente se, pelo menos, alguma coisa necessária para sua ação tem origem na própria agente. A ação não pode estar localizada em tempos e lugares anteriores ao livre desejo de agir da agente. Se uma agente não é a origem última de suas ações, então suas ações não se originam nela e, se suas ações são o resultado de condições que as garantem, como se pode dizer que a agente as controla? As condições suficientes para a ocorrência das ações já estavam presentes antes de a agente sequer existir!

Segue um argumento incompatibilista que demonstra as considerações discutidas acima:

A Uma agente qualquer, **x**, executa uma ação, **a**, de acordo com o seu livre-arbítrio se, e somente se, **x** tem controle sobre **a**.

B **x** tem controle sobre *a* somente se *x* é a origem última de *a*.

C Se *x* é a origem última de *a*, então alguma condição, *b*, necessária para *a*, é originada em *x*.

D Se qualquer condição, *b,* se origina em *x*, então não há qualquer condição suficiente para *b* independente de *x*.

E Se o determinismo for verdadeiro, então os fatos do passado em conjunção com as leis da natureza implicam todos os fatos sobre o futuro.

F Se os fatos do passado em conjunção com as leis da natureza implicam todos os fatos sobre o futuro, então, para qualquer condição, *b,* necessária para uma ação, *a*, executada por uma agente, *x*, existem condições independentes de *x* (no passado remoto de *x*, antes do nascimento de *x*) que são suficientes para *b*.

G Se, para qualquer condição, *b*, necessária para uma ação, *a*, executada por uma agente, *x*, existem condições independentes de *x* que são suficientes para *b*, então nenhuma agente, *x,* é a origem última de qualquer ação, *a* (segue-se de C e D).

H Se o determinismo for verdadeiro, então nenhuma agente, *x*, é a origem última de nenhuma ação, *a* (segue-se de E, F e G).

I Portanto, se o determinismo for verdadeiro, então nenhuma agente, *x*, executada qualquer ação, *a,* de acordo com o seu livre-arbítrio (segue-se de A, B e H).

Para facilitar a referência e a discussão ao longo deste verbete, vamos simplificar o argumento acima da seguinte maneira:

1 Uma pessoa age de acordo com o seu livre-arbítrio somente se ela é a origem última de suas ações (A-B).
2 Se o determinismo for verdadeiro, então ninguém é a origem última de suas ações (C-H).
3 Portanto, se o determinismo for verdadeiro, ninguém age de acordo com o seu próprio livre-arbítrio (I)

Chamaremos esse argumento simplificado de **Argumento a favor do Incompatibilismo de Origem**. É importante notar que a demanda por possibilidades alternativas não é, ao menos não obviamente, relevante para esse argumento incompatibilista. Suponha que uma agente com o livre-arbítrio tivesse acesso ao tipo relevante de possibilidades alternativas. De acordo com o Argumento a favor do Incompatibilismo de Origem, uma **condição adicional** é que ela deve ser a origem última de suas ações livres, isto é, o defensor ou a defensora da teoria da origem nega que as possibilidades alternativas sejam suficientes para garantir a liberdade.

1.3. Respostas Compatibilistas

Em resposta a esses argumentos, as compatibilistas e os compatibilistas negam que a liberdade requer a habilidade de agir de outro modo, que o determinismo causal impede a habilidade de agir de outro modo e, também, que a liberdade ou controle requer uma origem última. Nesse sentido, as(os) compatibilistas raramente negam a Premissa 2 do Argumento a favor do Incompatibilismo de Origem. Para exceções importantes, consulte McKenna (2008) e Bjornsson e Persson (2012). Mas as negações compatibilistas dessas premissas não são simplórias. Pelo contrário, elas são fundadas em concepções atrativas de agência humana. Então, em sequência, analisaremos os detalhes dessas teorias e, ao fazermos isso, descobriremos como diferentes teorias compatibilistas respondem aos argumentos incompatibilistas apresentados anteriormente.

2. Compatibilismo Clássico

A posição das compatibilistas e dos compatibilistas na filosofia contemporânea se desenvolveu em pelo menos três estágios. O primeiro estágio envolve a forma clássica de compatibilismo, que foi desenvolvida na era moderna pelos empiristas Hobbes e Hume e foi revigorada no início do século XX. O segundo estágio envolve três contribuições diferentes na década de 1960, as quais desafiaram muitos os pressupostos dialéticos que motivam o compatibilismo clássico. O terceiro estágio envolve várias formas contemporâneas de compatibilismo, que divergem da vertente clássica e que emergem de ao menos uma das três contribuições encontradas no segundo estágio. Esta seção é dedicada ao primeiro estágio, o do compatibilismo clássico.

2.1. Liberdade de acordo com o Compatibilismo Clássico

De acordo com uma vertente do compatibilismo clássico, a liberdade não é nada mais do que a habilidade de uma agente de fazer o que quer na ausência de impedimentos que poderiam obstruir seu caminho. Por exemplo, Hobbes (1997, p. 108) oferece uma expressão exemplar do compatibilismo clássico ao afirmar que a liberdade de uma pessoa consiste no fato de a pessoa "não encontrar nenhum impedimento ao fazer o que tem vontade, desejo ou inclinação para fazer." Nessa perspectiva, a liberdade envolve dois componentes, um positivo e um negativo. O componente positivo (fazer o que se tem vontade, desejo ou inclinação para fazer) consiste em nada mais do que aquilo que está envolvido no poder da agência. Já o componente negativo (não encontrar impedimentos) consiste em agir de maneira não forçada ou desimpedida. Tipicamente, a ação coagida é o padrão de referência do compatibilismo clássico sobre a ação impedida ou forçada. A ação coagida surge quando alguém é foçado por alguma força externa a agir de maneira contrária à própria vontade.

Para a compatibilista clássica e o compatibilista clássico, então, o livre-arbítrio é a habilidade de fazer o que se deseja. Portanto, é plausível concluir que a verdade do determinismo não implica que agentes não têm livre-arbítrio, dado que não envolve que agentes nunca fazem o que desejam fazer, nem que agentes são necessariamente forçadas ao agir. O compatibilismo é, portanto, reivindicado.

Mas o quão convincente é a posição do compatibilismo clássico sobre o livre-arbítrio? Tal como ela se apresenta, é necessário algum refinamento. Para citar apenas uma falha na teoria, várias doenças mentais podem fazer com que uma pessoa aja como ela deseja e faça isso sem ser forçada a tal, porém, intuitivamente, parece que a pessoa não agiria de acordo com o seu livre-arbítrio. Por exemplo, imagine uma pessoa que sofre de uma forma de psicose que causa alucinações completas. Durante as alucinações, ela pode "agir como ela deseja sem ser forçada", mas dificilmente se pode dizer que ela está agindo de acordo com o seu livre-arbítrio. Consequentemente, a compatibilista clássica e o compatibilista clássico nos devem oferecer mais explicações sobre tal posição. Para ver como pode ser complementada, voltaremos nossa atenção para uma maneira incompatibilista diferente na tentativa de eliminar o compatibilismo clássico.

2.2 A Análise Condicional Compatibilista Clássica

Considere a seguinte objeção incompatibilista à posição compatibilista clássica sobre o livre-arbítrio. Caso o determinismo seja verdadeiro e, se em algum momento, uma agente desimpedida for completamente determinada a ter os desejos que tem e, se aqueles desejos determinam causalmente as suas ações, então, mesmo que a agente realmente faça o que deseja fazer, **ela nunca pode agir de outro modo**. A agente satisfaz as condições compatibilistas clássicas de livre-arbítrio. Porém, o livre-arbítrio requer a habilidade de agir de outro modo, a qual é incompatível com o determinismo. Disso se segue que a posição compatibilista clássica sobre o livre-arbítrio é inadequada. O determinismo é incompatível com o livre-arbítrio e com a responsabilidade moral, visto que ele é incompatível com a habilidade de agir de outro modo.

A posição compatibilista clássica sobre a liberdade apresentada até agora pode ser considerada como uma defesa de uma liberdade **unidirecional**, que se fixa apenas no que uma pessoa faz, não nas alternativas que ela tinha ao fazê-lo. O desafio incompatibilista em questão é que essa liberdade, mesmo que necessária, é insuficiente na ausência de outra liberdade, a de se agir de modo diferente do qual se age.

As compatibilistas clássicas e os compatibilistas clássicos responderam argumentando que o determinismo *é* compatível com a habilidade de agir de outro modo. Para mostrar isso, tentaram analisar a habilidade de uma agente agir de outro modo em termos condicionais. A esse respeito, consulte Hume, na obra *Enquiry*

Concerning Human Understanding (1975, p. 73), Ayer (1954) ou Hobart (1934). Como o determinismo é uma tese sobre o que deve acontecer no futuro **dado o passado atual**, o determinismo é consistente com o futuro sendo diferente **dado um passado diferente**. Assim, as compatibilistas clássicas e os compatibilistas clássicos analisaram qualquer asserção de que uma agente poderia ter agido de outro modo como uma afirmação condicional relatando o que esta pessoa teria feito sob certas condições contrafactuais. Essas condições envolviam variações no que uma agente livre desejava, escolheu, quis ou decidiu fazer no momento de sua ação livre. Suponha que uma agente desejasse **X** livremente. De acordo com a **análise condicional** compatibilista clássica, dizer que, no momento da ação, a agente poderia ter feito **Y**, e não **X**, é apenas dizer que, se ela **desejasse**, **escolhesse**, **quisesse** ou **decidisse** fazer **Y** naquele momento, **então** ela teria feito **Y** e não **X**. Sua capacidade de agir de outro modo no momento em que agia consiste em uma verdade contrafactual desse tipo.

Porém, dado que uma agente específica está determinada, no momento da ação, a ter os desejos que tem, qual a utilidade em dizer o que ela teria feito se tivesse desejos diferentes dos que tinha? Assumindo a verdade do determinismo, no momento da ação, a agente não poderia ter tido outros desejos a não ser aqueles que sua história causal determinou que os tivesse.

Em resposta, a compatibilista clássica e o compatibilista clássico defendem que a análise condicional evidencia uma imagem rica da liberdade. Ao acessar a ação da agente, a análise distingue de maneira precisa aquelas ações que ela teria executado se desejasse, daquelas que ela não poderia ter executado mesmo que quisesse. A compatibilista clássica e o compatibilista clássico defendem que isso efetivamente distingue aqueles cursos de ação alternativos que estão no escopo da habilidade da agente no momento da ação, daqueles cursos de ação que não estão, e é a distinção entre o que uma agente era livre para fazer e o que ela não era livre para fazer. Isso não é de maneira alguma uma liberdade superficial, pois demarca o que as pessoas têm dentro de seu controle e o que está fora de alcance.

Apesar da ingenuidade compatibilista clássica, a análise de **poderia ter agido de outro modo** falhou decisivamente. As compatibilistas clássicas e os compatibilistas clássicos queriam mostrar às suas interlocutoras e aos seus interlocutores incompatibilistas que, quando é afirmado que uma agente livre tinha alternativas disponíveis, isto é, quando afirmado que ela poderia ter agido de outro modo, aquela afirmação poderia ser analisada como um enunciado condicional, perspicazmente compatível com o determinismo. Mas, no fim, a análise foi refutada

quando se mostrou que enunciados condicionais, às vezes, produziam o resultado indesejado de que uma pessoa era capaz de agir de outro modo mesmo sendo claro que, no momento em que a pessoa agiu, ela não tinha tais alternativas e que, portanto, não era capaz de agir de outro modo no sentido relevante (CHISHOLM, 1982, p. 26-27; VAN INWAGEN, 1983, p. 114-199). Vejamos o exemplo que se segue.

Suponha que Daniela seja psicologicamente incapaz de tocar um cachorro de pelo amarelado. Imagine que, em seu aniversário de 16 anos, sem saber dessa condição, o pai de Daniela traz dois filhotes de cachorro para que ela escolha entre um deles como presente, um dos cães tem pelo amarelado e o outro tem a pelugem preta. Ele diz a Daniela para pegar qualquer um dos dois que gostasse e que o outro filhote seria devolvido ao *pet shop*. A adolescente, muito feliz, e sem nenhuma coerção, faz o que ela quer e pega o filhote de pelo preto.

Quando Daniela pegou o filhote de pelo preto, ela era capaz de pegar o filhote de pelo amarelado? Parece que não. Pegar o filhote de pelo amarelado não era uma alternativa disponível para ela. Nesse sentido, ela **não poderia ter agido de outro modo**. Dada sua condição psicológica, ela não poderia nem mesmo formar o desejo de tocar o filhote de pelo amarelado, portanto, ela não poderia pegá-lo. Mas note que **se** ela desejasse pegar o filhote de pelo amarelado, **então ela o teria feito**. Claro que, se ela desejasse pegar o filhote de pelo amarelado, então ela não sofreria da desordem psicológica que causa sua incapacidade de tocar cachorros de pelo amarelado. Então, a análise compatibilista clássica de **poderia ter agido de outro modo** falha. De acordo com a análise, quando Daniela pegou o filhote de pelo preto, ela **era** capaz de pegar o de pelo amarelado, mesmo que, por causa da sua condição psicológica, ela **não fosse** capaz de agir de tal forma no sentido relevante. Portanto, a análise produz o resultado incorreto.

Logo, mesmo que uma agente não coagida faça o que deseja, se ela é determinada, ao menos como a incompatibilista e o incompatibilista sustentam, então a agente não poderia ter agido de outro modo. Segundo a objeção, já que o livre-arbítrio exige liberdade envolvendo possibilidades alternativas, a análise compatibilista clássica de liberdade falha.

2.3. A Influência Duradoura da Análise Condicional

O compatibilismo clássico falhou em responder ao Argumento Incompatibilista Clássico. O que tentaram fazer com a análise condicional foi negar a tese de que,

se o determinismo for verdadeiro, então, ninguém poderia agir de outro modo. Mas, dado o fracasso da análise, o compatibilismo deveria responder ao argumento de alguma maneira. É apenas dialeticamente justo reconhecer que o determinismo representa uma ameaça *prima facie* ao livre-arbítrio, entendido como aquele que exige a habilidade de agir de outro modo. O Argumento Incompatibilista Clássico é meramente uma expressão desse pensamento comum. À luz do fracasso da tradicional análise condicional compatibilista, o ônus da prova está totalmente com as compatibilistas e os compatibilistas. De que maneira a habilidade de agir de outro modo pode ser reconciliada com o determinismo? Como veremos a seguir, as(os) compatibilistas atuais tentaram responder esse problema.

3. Compatibilismo em Transição

Na década de 1960, três grandes contribuições ao debate sobre o livre-arbítrio o alteraram radicalmente. Uma delas foi um argumento incompatibilista que, claramente, capturou a intuição de que uma agente determinada não tem controle sobre alternativas. Esse argumento, desenvolvido pela primeira vez por Carl Ginet, passou a ser conhecido como o **Argumento da Consequência** (GINET, 1966). Outra contribuição foi o argumento de Harry Frankfurt (1969) contra o **Princípio das Possibilidades Alternativas** (PPA), que afirma que uma agente é moralmente responsável pelo que faz apenas se puder agir de outro modo. Finalmente, P.F. Strawson (1962) defendeu o compatibilismo convidando filósofas(os) compatibilistas e incompatibilistas a lidarem, de maneira mais cuidadosa, com o papel central dos relacionamentos interpessoais e das **atitudes reativas** na compreensão do conceito de responsabilidade moral. Cada uma dessas contribuições mudou drasticamente a maneira como o problema do livre-arbítrio é tratado nas discussões contemporâneas. Hoje em dia, nenhuma teoria sobre o livre-arbítrio, compatibilista ou incompatibilista, é proposta sem levar em consideração pelo menos uma, se não mais, dessas três grandes contribuições.

3.1. O Argumento da Consequência

Esse argumento envolve um padrão de inferência convincente, que talvez esteja oculto no Argumento Incompatibilista Clássico, a respeito das teses sobre a

necessidade agencial. A necessidade agencial, aplicada a proposições verdadeiras, ou fatos, diz respeito ao que **não** está no escopo do controle de uma pessoa, ou seja, diz respeito a fatos sobre os quais uma pessoa não tem controle. **Dizer que uma pessoa não tem controle sobre um fato é dizer que ela não é capaz de agir de tal modo que o fato não aconteceria**. Para ilustrar, nenhuma pessoa tem controle sobre as verdades matemáticas. Isto é, nenhuma pessoa é capaz de agir de tal modo que as verdades matemáticas fossem falsas. Logo, as verdades matemáticas são, para qualquer pessoa, agencialmente necessárias.

O padrão de inferência intuitivo aplicado a essas teses é o seguinte: se uma pessoa não tem controle sobre certo fato e, se ela também não tem controle sobre o fato adicional de que o original tem outro como consequência, então ela também não tem controle sobre o fato consequente. Parece que a ausência de controle é transferida de um fato para as suas consequências. Por exemplo, se um jogador de pôquer, Joãozinho, que possui apenas dois pares, não tem controle sobre o fato de que Samuel tira um *straight flush* e, se um *straight flush* vence dois pares, e assumindo que Joãozinho não é capaz de alterar esse fato, então, se segue que Joãozinho não tem controle sobre o fato de que a mão de Samuel vence a de Joãozinho. Esse padrão geral de inferência é aplicado à tese do determinismo causal para produzir um argumento poderoso a favor do incompatibilismo. O argumento exige a suposição que o determinismo é verdadeiro e que os fatos sobre o passado e as leis da natureza são fixos. Dadas essas suposições, eis um esboço não-técnico do argumento:

1 Ninguém tem controle sobre os fatos acerca do passado e das leis da natureza.
2 Ninguém tem controle sobre o fato de que fatos acerca do passado e as leis da natureza implicam todos os fatos acerca do futuro (isto é, o determinismo é verdadeiro).
3 Portanto, ninguém tem controle sobre os fatos acerca do futuro.

De acordo com o Argumento da Consequência, **se** o determinismo for verdadeiro, parece que ninguém tem controle algum sobre como o seu próprio futuro se desdobrará.

Esse argumento chocou as(os) compatibilistas, e com razão. O fracasso do compatibilismo clássico em analisar enunciados sobre a capacidade de uma agente em termos de condicionais contrafactuais deixou as compatibilistas e os compatibilistas sem nenhuma resposta perspicaz contra a segunda premissa do

Argumento a favor do Incompatibilismo Clássico. E o Argumento da Consequência fornece um suporte poderoso para essa segunda premissa do argumento. Se, de acordo com o Argumento, o determinismo implica que o futuro só se desdobrará de uma única maneira, dado o passado e as leis da natureza, e se ninguém tem controle algum para alterar esse desdobramento dessa maneira particular, então parece que ninguém é capaz de agir de maneira diferente da que age.

É justo dizer que o Argumento da Consequência concedeu às(aos) incompatibilistas uma vantagem dialética importante. O compatibilismo nos deve uma explicação sobre o que há de errado no Argumento da Consequência e, talvez, alguma explicação positiva sobre a habilidade de agir de outro modo. Então, mesmo que muitas compatibilistas e muitos compatibilistas estejam comprometidos com a tese de que o Argumento da Consequência não é correto, o argumento, mesmo assim, definiu a agenda para muitas teorias compatibilistas contemporâneas sobre o livre-arbítrio e a responsabilidade moral.

3.2. Um Desafio ao Princípio de Possibilidades Alternativas

Uma estratégia de resposta compatibilista ao Argumento a favor do Incompatibilismo Clássico consiste em conceder que, talvez, o Argumento da Consequência nos forneça uma boa razão para pensar que o determinismo elimina a habilidade de agir de outro modo, ao mesmo tempo em qu defende que tal capacidade não é necessária para o livre-arbítrio. Em outras palavras, a compatibilista e o compatibilista podem contornar as questões levantadas pelo Argumento da Consequência atacando diretamente a primeira premissa do Argumento Incompatibilista Clássico, o qual afirma que se uma pessoa age de acordo com o seu livre-arbítrio, então, ela poderia ter agido de outro modo. Essa resposta compatibilista rejeita uma concepção da agência humana que situa o controle na capacidade de agir de outro modo. Como alternativa, procura fundamentar o controle de uma agente sobre sua ação em outros aspectos de sua agência. Em *Alternate Possibilities and Moral Responsibility* (1969), Harry Frankfurt desenvolve um argumento poderoso que fornece os recursos exigidos para discutir exatamente dessa maneira. O argumento de Frankfurt foi direcionado contra o Princípio das Possibilidades Alternativas (PPA):

PPA: Uma pessoa é moralmente responsável pelo que fez somente se ela é capaz de agir de outro modo.

O que é central no ataque de Frankfurt a PPA é uma série de exemplos nos quais uma agente é moralmente responsável por sua conduta, mas não é capaz, no momento da ação em questão, de agir de outro modo. Eis uma versão aproximada do exemplo apresentado por Frankfurt em seu artigo original:

Jones decidiu que vai atirar em Smith. Black aprendeu o plano de Jones e quer que Jones atire em Smith. Mas Black prefere que Jones atire em Smith por conta própria. No entanto, preocupado que Jones hesite em sua decisão de atirar em Smith, Black organiza as coisas em segredo para que, se Jones mostrar qualquer sinal de que não atirará em Smith, algo que Black tem recursos para detectar, Black poderá manipular Jones de tal maneira que ele atire em Smith. Mas, no decorrer da situação, Jones segue os seus planos e atira em Smith por seus próprios motivos. Ninguém, da maneira que seja, ameaçou ou coagiu Jones, ofereceu-lhe suborno ou sugeriu que atirasse em Smith. Jones atirou em Smith por sua própria vontade. Black nunca interveio.

Nesse exemplo, Jones atirou em Smith por vontade própria e sem coerção, ou seja, fez isso livremente. Mas, dada a presença de Black no cenário, Jones não poderia ter falhado em atirar em Smith, isto é, ele não era capaz de agir de outro modo. Logo, temos um contraexemplo ao PPA.

Se o argumento de Frankfurt contra o PPA for correto, então o debate sobre o livre-arbítrio foi sistematicamente mal interpretado na maior parte da história da Filosofia. *Se* o determinismo ameaça o livre-arbítrio e a responsabilidade moral, **não** é por causa de sua incompatibilidade com a habilidade de agir de outro modo. Mesmo que o determinismo **seja** incompatível com um tipo de liberdade envolvendo a habilidade de agir de outro modo, **esse não é o tipo de liberdade requerida para a responsabilidade moral**.

Talvez não surpreenda que tenha surgido uma literatura enorme, e complexa, sobre o sucesso do argumento de Frankfurt e, em particular, sobre o exemplo que Frankfurt ofereceu contra o PPA. O debate está bastante vivo e ainda não temos um claro vencedor, de forma que incompatibilistas possam afirmar, com razão, que acabaram definitivamente com a estratégia da análise condicional de compatibilistas (*vide* seção 3.3). Independentemente disso, o que é mais relevante para este ensaio é que o argumento de Frankfurt inspirou muitas e muitos compatibilistas a começarem a pensar em explicações sobre a liberdade ou o controle que se afastam, de maneira evidente, das possibilidades alternativas.

3.3. O Foco nas Atitudes Reativas

No texto *Freedom and Resentment* (1962), P. F. Strawson rompeu com o compatibilismo clássico. Srawson desenvolveu três argumentos distintos a favor do compatibilismo, argumentos bastante diferentes daqueles defendidos pelas compatibilistas clássicas e pelos compatibilistas clássicos. Porém, mais relevante do que seus argumentos foi sua teoria geral sobre o que é a responsabilidade moral e, portanto, sobre o que está em jogo ao argumentar sobre ela. Strawson defende que tanto as(os) incompatibilistas quanto as(os) compatibilistas interpretaram erroneamente a natureza da responsabilidade moral. Strawson sugere que, ambos os membros dessa disputa, avançaram em argumentos a favor ou contra uma imagem distorcida da realidade.

Para entender a responsabilidade moral da maneira adequada, Strawson convida a leitora e o leitor a considerarem as **atitudes reativas** que alguém tem, em relação a outra pessoa, quando reconhece na conduta da outra pessoa uma atitude de má vontade. As reações que naturalmente se seguem de testemunhar má vontade são, elas mesmas, atitudes que são direcionadas às intenções ou atitudes do ofensor ou da ofensora. Quando um ofensor ou uma ofensora faz um mal a uma pessoa, ela, a parte que sofre a ofensa, normalmente tem uma atitude pessoal reativa de ressentimento. Quando o ofensor ou a ofensora faz um mal à outra pessoa, uma terceira parte que sofre a ofensa, a atitude reativa natural é a indignação moral, que equivale a um caso análogo do ressentimento sentido em nome da parte que sofre a ofensa. Quando alguém é a parte que ofende, ao refletir sobre isso e perceber que ofendeu a outrém, a atitude reativa natural é a culpa.

Strawson queria que as e os participantes do debate do livre-arbítrio vissem com mais clareza que desculpar uma pessoa, escolhendo não a considerar merecedora de culpa, envolve mais do que um julgamento objetivo de que ela não fez tais e tais coisas, ou não tinha essa ou aquela intenção e, portanto, não merece esse ou outro tratamento específico. Tal atitude inclui uma suspensão ou retirada de certas práticas moralmente reativas que envolvem respostas emocionais. Na perspectiva de Strawson, responsabilizar moralmente uma pessoa por sua conduta **não é** nada mais do que a propensão em relação a (ou a sustentação de) uma atitude moralmente reativa, como a indignação. A indignação é, crucialmente, uma resposta à atitude percebida de má vontade ou ao motivo culpável na conduta da

pessoa responsabilizada. Logo, Strawson explica: colocar em questão se todo o arcabouço da responsabilidade moral deve ser considerado irracional (se fosse descoberto que o determinismo é verdadeiro) equivale a colocar em questão se as pessoas inseridas em uma comunidade interpessoal, isto é, na vida real, deveriam deixar de ter as atitudes reativas em relação às pessoas que agem errado com outras e que, às vezes, fazem isso intencionalmente. Strawson nos convida a notar que as atitudes moralmente reativas, que são as bases constitutivas de nossas práticas de responsabilidade moral, bem como as relações interpessoais e as expectativas que estruturam tais atitudes, são profundamente interligadas à vida humana. Essas atitudes, relações e expectativas são uma expressão de características básicas e naturais de nossas vidas sociais, de suas texturas emocionais, tanto que é praticamente inconcebível imaginar como elas **poderiam ser** descartadas.

4. Compatibilismo Contemporâneo

Três grandes contribuições na década de 1960 alteraram profundamente a face do compatibilismo: o Argumento Incompatibilista da Consequência, o ataque de Frankfurt ao Princípio das Possibilidades Alternativas (PPA) e o foco de Strawson nas atitudes reativas. Qualquer posição compatibilista resultante na literatura contemporânea é moldada de alguma maneira por pelo menos uma dessas influências. Esta seção se concentrará em seis das posições compatibilistas contemporâneas mais importantes. Aquelas e aqueles que desejam conhecer os trabalhos de ponta sobre o assunto podem ler o complemento **Compatibilismo: o Estado da Arte**[3] .

4.1. Compatibilismo com relação à Liberdade para Agir de Outro Modo

O Argumento da Consequência (seção 3.1) é uma forte defesa a favor da incompatibilidade entre o determinismo e a liberdade para agir de outro modo.

---

[3] MCKENNA,M.; COATES, D. J. Supplement to Compatibilism. Compatibilism: State of the Art. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Summer Edition. Stanford, CA: The Metaphysics Research Lab, 2020. Disponível em: https://plato.stanford.edu/archives/sum2020/entries/compatibilism/supplement.html. Acesso em: 20 ago. 2021.

Assumindo que o determinismo é verdadeiro, ele pressupõe que:

1 Ninguém tem controle sobre os fatos acerca do passado e das leis da natureza.
2 Ninguém tem controle sobre o fato de que fatos acerca do passado e as leis da natureza implicam todos os fatos acerca do futuro (isto é, o determinismo é verdadeiro).
3 Portanto, ninguém tem controle sobre os fatos acerca do futuro.

As(os) compatibilistas que aceitam que as possibilidades alternativas são necessárias para a responsabilidade moral devem mostrar o que há de errado com esse argumento poderoso. Também deveriam oferecer uma explicação para o que John Martin Fischer (1994) chamou de **controle regulador**, ou seja, uma forma de controle que agentes possuem quando são capazes de fazer com que **X** seja o caso e capazes de escolher com que **X** não seja o caso, que deixa claro como isso é possível mesmo em um mundo determinístico. Primeiro, consideraremos três diferentes tentativas compatibilistas de derrotar o Argumento da Consequência. Em seguida, veremos como algumas e alguns compatibilistas, as chamadas "Novas disposicionalistas" ou os chamados "Novos disposicionalistas", explicam o controle regulador, ou seja, como podem explicar a liberdade de agir de outro modo de uma maneira compatível com o determinismo causal.

4.1.1. Desafiando o Controle Necessário e o Passado

Algumas e alguns compatibilistas argumentaram contra a primeira premissa do Argumento da Consequência tentando mostrar que uma pessoa **é capaz** de agir de tal modo que o passado seria diferente. Considere a diferença entre uma pessoa no presente, que tem a capacidade de agir de tal maneira que **ela altera o passado**, em oposição a uma pessoa que tem a capacidade de agir de tal maneira que, **se ela agisse**, **então o passado seria diferente**. Note que a primeira capacidade é inusitada, pois ela exigiria poderes mágicos. No entanto, a segunda capacidade, ao menos por comparação, é incontroversa, já que ela apenas indica que uma pessoa que agiu de certa forma, em certo momento, possuía a capacidade de agir de diversas formas. Se ela tivesse exercido uma dessas capacidades e, por conta disso, tivesse agido de maneira diferente, então, o passado que antecedeu sua ação teria sido diferente. Para ilustrar o quão comparativamente fraca pode ser

essa tese sobre uma capacidade de uma agente e o passado, pense no tipo logicamente similar de tese que é, simplesmente, acerca do que seria exigido para uma agente agir de maneira diferente. Por exemplo, considere a afirmação: **se eu estivesse dançando na Riviera Francesa agora, eu seria muito mais rico do que eu sou**. Certamente, essa afirmação não significa que, se eu for dançar na Riviera Francesa, então eu **ficarei** mais rico. Significa apenas que, se eu tivesse ido para lá dançar tango, então, eu teria que ter muito mais dinheiro de antemão para financiar minhas aventuras. Algumas (alguns) compatibilistas (*vide* SAUNDERS, 1968; PERRY, 2010; DORR, 2016) alegaram que a defesa do Argumento da Consequência se apoia na noção excêntrica de capacidade na primeira premissa do argumento. Mas, segundo tais compatibilistas, a primeira premissa é falseada quando interpretada com uma noção mais fraca de capacidade.

4.1.2. Desafiando o Controle Agencial e as Leis da Natureza

Outras(os) compatibilistas argumentaram contra a primeira premissa do Argumento da Consequência de uma maneira paralela, tentando mostrar que uma pessoa **pode** agir de tal maneira que uma lei da natureza não existiria. Assim como a distinção feita a respeito da capacidade e do passado, considere a diferença entre uma pessoa que tem a capacidade de agir de tal maneira que **ela viola uma lei da natureza**, em oposição a uma pessoa, em um mundo determinístico, que tem a capacidade de agir de tal maneira que, **se ela agisse**, **então alguma lei da natureza que existe não existiria**. Note que a primeira capacidade exigiria poderes mágicos. Já a segunda, de acordo com as(os) compatibilistas, não exigiria nada estranho, apenas nos diz que uma pessoa que agiu de certa maneira e, em certo momento, possuía capacidades para agir de várias maneiras distintas. Se a pessoa tivesse exercido uma dessas capacidades, e agido de maneira diferente, então, as leis da natureza que teriam implicado o que a pessoa fez, naquele cenário hipotético, seriam diferentes das atuais leis da natureza, que implicaram o que a pessoa realmente fez. Essa última capacidade não assume que agentes são capazes de violar leis da natureza, mas apenas assume que, quaisquer que sejam as leis da natureza (ao menos em mundos determinísticos), elas precisam ser tais que implicam, dado o passado, o que uma agente vai fazer. Se uma agente atua de maneira diferente, em algum mundo possível ao invés do atual, então, algum outro conjunto de leis será o que implica o que a agente faz naquele mundo. Algumas e alguns

compatibilistas (LEWIS, 1981; GRAHAM, 2008; PENDERGRAFT, 2011), fixando-se na capacidade pertinente às leis da natureza, justificaram que o Argumento da Consequência se apoia na estranha noção de capacidade na primeira premissa do argumento. Mas, segundo tais compatibilistas, a primeira premissa é falseada quando interpretada com uma noção incontroversa de capacidade.

4.1.3. Desafiando as Inferências baseadas na Necessidade Agencial

Michael Slote (1982) tentou refutar o Argumento da Consequência, mostrando que sua inferência central é inválida. O ponto central, em torno do qual Slote trabalha, é que noções como **inevitabilidade**, ou **necessidade agencial**, são sensíveis a contextos, de modo que permitem apenas "seletivamente" o tipo de inferência em questão no Argumento da Consequência. Trabalhemos com a ideia de **inevitabilidade**. De acordo com Slote, quando dizemos que algo é inevitável para uma pessoa, temos em mente contextos "seletivos", nos quais os fatos pertencentes à inevitabilidade não têm nada a ver com aquela pessoa, pois eles ignoram completamente a agência da pessoa (SLOTE, 1982, p. 19). É inevitável para mim, por exemplo, que César atravessou o rio Rubicão, ou que a maioria dos veículos motorizados funcione com gasolina. Nada sobre a minha agência, sobre o que sou capaz de fazer, pode alterar tais fatos. Isso sugere que a inevitabilidade é aplicada incorretamente quando se trata de aspectos da agência de uma pessoa. Mas note que, no Argumento da Consequência, a **inevitabilidade**, ou **necessidade agencial**, passa de um contexto no qual a noção é aplicada adequadamente a um contexto que, segundo Slote, ela não é empregada de forma correta. A primeira premissa trata de coisas que não têm nada a ver com a agência de uma pessoa, fatos anteriores ao seu nascimento e às leis da natureza. Alega-se que esses fatos são inevitáveis para uma pessoa, mas, a partir disso, conclui-se que as próprias ações executadas por uma pessoa são inevitáveis para ela. No entanto, Slote e outras(outros) compatibilistas (DENNETT, 1984a; MCKAY; JOHNSON, 1996) sugeriram que fazer isso é tirar, de maneira ilícita, conclusões incompatibilistas de alegações razoáveis sobre inevitabilidade.

4.1.4. Dando uma Explicação sobre a Liberdade para Agir de Outro Modo

Mesmo que algumas respostas compatibilistas provem que o Argumento da Consequência não é correto, isso por si só não seria um argumento positivo a favor do compatibilismo. Apenas significaria que um dos principais argumentos a favor da incompatibilidade, entre o determinismo e o controle regulador, é insustentável. Mas isso é consistente com a **incompatibilidade** do determinismo e do controle regulador. De fato, há defensoras e defensores dessa incompatibilidade que não confiam nas suposições potencialmente problemáticas acerca da necessidade agencial em jogo no Argumento da Consequência (FISCHER, 1994; GINET, 1990, 2003). Além disso, mesmo que a(o) compatibilista pudesse refutar todos os argumentos atuais a favor da incompatibilidade entre o determinismo e o controle regulador, ainda lhe cabe oferecer um argumento positivo que demonstre a compatibilidade entre o determinismo e o controle regulador. As(os) compatibilistas que desejam defender o controle regulador, como Berofsky (1987, 1995, 2012), Campbell (1997), Nelkin (2011) e Vihvelin (2013), ainda têm um trabalho difícil a ser feito.

4.1.5. O Novo Disposicionalismo

Recentemente, o compatibilismo ofereceu uma explicação positiva do controle regulador (*vide* FARA, 2008; M. SMITH, 2003; VIHVELIN, 2004, 2013). Chamaremos a posição proposta pelo compatibilismo de **novo disposicionalismo**.

Ao propor uma tese compatibilista, Vihvelin (2004, p. 429) trata da habilidade de agir de outro modo, e especialmente de escolher diferentemente, em termos de um grupo de disposições. Da mesma forma, Fara propõe uma análise disposicional da capacidade de agir de outro modo. Já Smith (2003, p. 27) entende as capacidades racionais de crer e desejar de outro modo, e, por consequência, agir de outro modo, em termos de um "amontoado de possibilidades". Para Fara, Vihvelin e Smith, acessamos afirmações sobre a disposição constitutiva da habilidade de agir de outro modo, ou das disposições no pacote, ou ainda das possibilidades no amontoado, atentando-nos às propriedades intrínsecas de uma agente em virtude da qual ela age quando tenta agir (FARA, 2008, p. 861), ou à base causal das disposições relevantes (VIHVELIN 2004, p. 436), ou à estrutura fundamental de uma capacidade racional (SMITH, p. 29). Mas como? Fara não diz, mas parece que o autor concordaria com algo próximo à proposta oferecida por Vihvelin e Smith. Segundo esses últimos

autores, mantemos a base causal relevante fixa, ou a estrutura fundamental, da disposição de uma agente para, digamos, acenar ao cumprimentar uma amiga ou um amigo ou dizer a verdade sob interrogatório, e consideramos várias condições contrafactuais nas quais a base causal, ou a estrutura fundamental, opera de maneira intacta. A agente, em uma rica variedade de condições contrafactuais, acena ou diz a verdade? Se sim, então mesmo que no mundo atual ela não acene ou diga a verdade, ela era capaz de o fazer. Ela tinha, no momento da ação, as habilidades ou capacidades agenciais relevantes. E isso é verdade mesmo que esse mundo seja determinístico (VIHVELIN, 2004, p. 437). Por quê? Porque não há base argumentativa quando testamos as disposições relevantes em outros mundos possíveis, já que temos que restringir os mundos àqueles nos quais mantemos fixos o passado e as leis. Observe como os problemas são contornados com a análise contrafactual do compatibilismo clássico. Se detivermos atenção à base causal das disposições relevantes, poderemos ver facilmente como a base causal relevante de Daniela está mal equipada para pegar cães com pelos amarelados.

O novo disposicionalismo, claramente, é uma melhora em relação ao compatibilismo clássico. Mas como ele se sai sozinho? Temos aqui uma explicação positiva convincente da habilidade, e da liberdade, de agir de outro modo que é compatível com o determinismo? Um ponto sinuoso tem a ver com a maneira como os mundos relevantes são identificados nos parágrafos anteriores. Temos que restringir nossa atenção aos mundos possíveis nos quais a base causal, ou a estrutura fundamental, da habilidade de agir de outro modo opera intacta. Algumas pessoas defenderão que essa restrição não é dialeticamente inocente. Considere um exemplo de Frankfurt (seção 3.2). Suponha que Jones atire livremente em Smith, mas se ele estivesse prestes a agir de outro modo, Black faria Jones agir contra a sua vontade (a vontade de Jones). Quando Jones atira em Smith por vontade própria, ele assim o faz de maneira livre e é moralmente responsável, apesar do fato de que, graças à presença de Black, ele não era livre para agir de outro modo. Fara (2008, p. 854-855), Nelkin (2012), Smith (2003, p. 19) e Vihvelin (2013, p. 445-448) dizem o contrário, pois defendem que Jones poderia ter agido de outro modo, era capaz de agir de forma diferente e era livre para o fazer quando atirou em Smith por vontade própria. Por quê? Grosso modo, porque se fixarmos na estrutura causal fundamental implicada pelo evento de que Jones atira em Smith por conta própria, e, se nos atentarmos apenas para os mundos possíveis, nos quais essa estrutura causal opera, sem impedimentos, então excluiremos os mundos nos quais o interventor contrafactual, Black, opera. Assim, seremos capazes de especificar uma

variedade de contrafactuais verdadeiras nas quais uma agente teve algum motivo, por exemplo, para agir de outro modo e, de fato, assim o fez. A questão delicada aqui, que não tentaremos responder, é se, **ao dar uma explicação para a liberdade de agir de outro modo**, as novas e os novos disposicionalistas têm o direito de restringir a atenção apenas aos mundos possíveis nos quais a base causal relevante opera sem impedimentos.

Outros compatibilistas, mais notadamente John Martin Fischer (1994) e Fischer e Mark Ravizza (1998), que podem ser consultados na seção 4.5, apelaram a restrições similares. Mas, ao fazer isso, os autores apenas pretendiam explicar a natureza da liberdade, ou controle, exibido na maneira como a agente agiu. Tal aspecto é o que Fischer chamou de **controle de direcionamento**. Em notável contraste de como as novas disposicionalistas pensam, Fischer e Ravizza não acham que têm o direito de afirmar que uma agente, no exemplo de Frankfurt, é livre para agir de outro modo. Então, é possível que o que o novo disposicionalismo identificou, com as contrafactuais em questão, não seja a liberdade para agir de outra forma, mas, sim, uma liberdade localizada no que uma agente de fato faz (o que é um assunto de controle de direcionamento, não de controle regulador). Pelo menos é desse jeito que compatibilistas, como Fischer e Ravizza, argumentariam. Para uma discussão movimentada sobre esse assunto, confira o debate entre Fischer (2008) e Vihvelin (2008).

4.2. Compatibilismo Hierárquico

Por trás de sua rejeição ao Princípio das Possibilidades Alternativas, Harry Frankfurt (1971) desenvolveu uma teoria compatibilista que não faz qualquer apelo ao controle regulador. Em vez disso, a **teoria do entrosamento hierárquico** pode ser vista como um desenvolvimento das tentativas compatibilistas clássicas de entender a liberdade em termos da capacidade desimpedida de uma agente conseguir o que quer (*vide* seção 4.1). Mais precisamente, Frankfurt explica que as ações livres são aquelas que resultam de desejos que combinam perfeitamente com elementos ordenados hierarquicamente em relação à psicologia de uma pessoa. A ideia principal é que uma pessoa, que age de acordo com seu livre-arbítrio, age a partir de desejos que estão ligados a elementos mais abrangentes do seu Eu. Nessa perspectiva, quando uma agente livre executa uma ação, suas ações emanam **dela mesma** e não de algo exterior.

4.2.1. Desejos de Ordem-Superior e Livre-Arbítrio

Frankfurt propõe uma distinção entre desejos de primeira-ordem e segunda-ordem. Isso serve como base para sua explicação hierárquica de liberdade. Os desejos de primeira-ordem têm como objeto ações, como comer uma fatia de bolo, assistir a um filme ou mexer os quadris ao som de Cartola. Já os desejos de segunda-ordem são desejos sobre desejos. Eles têm como objeto desejos de primeira-ordem, como o desejo de ter motivação para se exercitar diariamente, algo que, lamentavelmente, muita gente não tem: "Se ao menos eu **quisesse** ir à academia hoje, então seria fácil eu sair deste sofá!".

Entre os desejos de primeira-ordem que uma pessoa tem, alguns **não** movem a pessoa para a ação, como o desejo (insatisfeito) de dizer ao seu chefe o que sabe que não deve dizer. Outros desejos de primeira-ordem, no entanto, **movem** uma pessoa para agir, como o desejo (satisfeito) de cumprir o pedido do(a) chefe. Frankfurt (1971, p. 84) identifica a **vontade** de uma agente com seu desejo efetivo de primeira-ordem, aquele que move uma pessoa "por todo o caminho até à ação".

Frankfurt também distingue diferentes tipos de desejos de segunda-ordem. Alguns são meramente desejos de ter desejos de primeira-ordem, mas **não** que eles sejam inclusos na vontade da pessoa. Frankfurt usa o exemplo de um psicoterapeuta que quer experienciar o desejo por drogas para entender melhor um paciente. O terapeuta não tem desejo de que seja eficiente em levá-lo à ação (1971, p. 84-85). Ele quer saber como é sentir o desejo pela droga, mas não tem interesse em usá-la. Por outro lado, outros desejos de segunda-ordem que uma pessoa tem são por aqueles de primeira-ordem **efetivos**, que seriam inclusos na sua vontade e, assim, seriam eficazes em movê-la por todo o caminho até a ação. Por exemplo, uma pessoa em dieta, que é constantemente frustrada por seus desejos por açúcar, pode desejar ter mais saúde, o que seria mais eficiente em guiar seus hábitos alimentares do que costuma. Frankfurt chama esses desejos de segunda-ordem de **volições de segunda-ordem**. Não há limite teórico para o grau de ordem dos desejos. A pessoa em dieta pode desenvolver um desejo de terceira-ordem para que o seu de segunda-ordem, em relação à saúde, não desempenhe um papel tão dominante em suas deliberações diárias. Ela pode concluir que outras coisas são mais importantes na vida do que se preocupar com suas motivações alimentares.

Uma vez apresentado esse aparato conceitual, Frankfurt faz uma comparação entre diferentes tipos de viciados para ilustrar seu conceito de livre-arbítrio. Considere primeiro o **viciado involuntário**, que é alguém que tem ambos os desejos de primeira-ordem de usar a droga e de não usar a droga. No entanto, o viciado involuntário também tem uma volição de segunda-ordem de que o seu desejo de primeiro-ordem de usar a droga **não seja a sua vontade**. Essa é a base para considerarmos o viciado como involuntário. Porém, infelizmente, o seu desejo irresistível e viciante de usar a droga constitui a sua vontade. Em seguida, considere o caso do **viciado voluntário**. O viciado voluntário, assim como o involuntário, tem desejos de primeira-ordem conflitantes em relação ao uso da droga na qual é viciado. Mas, o viciado voluntário, por uma volição de segunda-ordem, aceita o seu desejo de primeira-ordem de usar a droga. Ele quer ser do jeito que é e quer agir do jeito que age.

Agora fica mais fácil ilustrar a teoria hierárquica de livre-arbítrio de Frankfurt. O viciado involuntário não usa a droga livremente, dado que sua vontade conflita em um grau superior com o que ele quer que a droga seja. Por sua vez, o viciado voluntário usa a droga livremente, pois a vontade dele se encaixa com o que ele quer que a droga seja. Assim, a teoria de Frankfurt pode ser expressa da seguinte maneira:

Uma pessoa age de acordo com o seu livre-arbítrio se, e somente se, sua ação resulta da vontade do que a pessoa deseja.

Pode parecer estranho que o viciado voluntário de Frankfurt aja livremente visto que, por seu vício, ele não é capaz de agir de outro modo. No entanto, é preciso lembrar que Frankfurt não acredita que a liberdade que envolve possibilidades alternativas seja requerida para a responsabilidade moral. Ao invés disso, Frankfurt acredita que a liberdade relevante para a responsabilidade moral diz respeito ao que uma agente faz e a seus motivos reais para fazê-lo. Ou seja, o autor acredita que, retornando à distinção útil de Fischer, é o controle de direcionamento que é necessário para a responsabilidade moral, não o controle regulador. O viciado voluntário possui o tipo de liberdade exigida para a responsabilidade moral porque a vontade que o leva à ação é a que ele deseja, agindo com base no controle de direcionamento.

4.2.2. Dois Problemas para a Teoria Hierárquica

A teoria hierárquica de Frankfurt enfrentou críticas minuciosas. Aqui, vamos considerar duas objeções que emergem de aspectos estruturais da teoria. Uma delas tem a ver com sua natureza hierárquica. A outra, por sua vez, com a sua dependência exclusiva em uma mistura entre diferentes características da psicologia de uma agente. Para uma discussão das tentativas de Frankfurt de responder a esses problemas, consulte a **seção A,** do complemento **Compatibilismo: o Estado da Arte**.

Considere o **problema hierárquico**. De acordo com Frankfurt, uma pessoa que enfrenta um problema que diz respeito à liberdade de sua vontade, está em conflito sobre um (ou mais) de seus desejos de primeira-ordem. Qual dos desejos "fala" pela agente? Com qual dos desejos ela pode ser identificada? Para resolver conflitos entre desejos de primeira-ordem, ou entre as atitudes de uma agente sobre um desejo de primeira-ordem específico, Frankfurt nos diz que uma pessoa forma um desejo de segunda-ordem que tem como objeto o(s) desejo(s) de primeira-ordem que ela quer que a mova no sentido da ação. Desse modo, a agente endossa ou se identifica com um dos desejos de primeira-ordem e, se tudo correr bem, esse desejo se torna sua vontade. Por esse processo, ela adiciona ao seu Eu um tipo de desejo e aliena o outro. Porém, aqui surge um problema, visto que uma pessoa não pode ser identificada com seus desejos de primeira-ordem porque ela pode se alienar a eles. Entretanto, ela também pode se alienar de desejos de segunda-ordem, ou de ordens-superiores (WATSON, 1975). No fim, a agente pode se encontrar tanto querendo que um desejo específico de primeira-ordem seja a sua vontade, quanto pode estar horrorizada com tal pensamento, ou seja, não querer que esse desejo de primeira-ordem em particular seja sua vontade. Logo, o problema do livre-arbítrio de uma agente pode reaparecer nesses estágios sempre em ascensão. Se isso estiver correto, a visão de Frankfurt está incompleta. Talvez sua posição sobre o livre-arbítrio, de fato, articule uma condição necessária para agir de acordo com o próprio livre-arbítrio, mas não parece ser o suficiente. Ela precisa ser complementada para evitar o problema de uma recorrência em espiral de desafios à liberdade de uma agente.

Agora considere o **problema do entrosamento**. De acordo com Frankfurt, se uma ação livre, pela qual uma agente é moralmente responsável, depende totalmente da relação entre a vontade da agente e suas volições de segunda-ordem, então, não importa a maneira pela qual a agente passa a ter esse entrosamento particular. Mas podemos construir alguns casos que parecem sugerir que **realmente**

importa a maneira pela qual a agente passa a ter um entrosamento particular entre seus desejos de primeira-ordem e os de segunda-ordem (*vide* SLOTE, 1980; FISCHER; RAVIZZA, 1998, p. 194-206). Usando o próprio exemplo de Frankfurt, do viciado voluntário, suponha que o desejo de segunda-ordem seja causado pelos efeitos do uso da droga e que ela tenha prejudicado os julgamentos, ou preferências, do usuário, oriundos de uma reflexão de segunda-ordem sobre seu próprio estado mental. Ou, ainda, deixando esse tipo de caso de lado, imagine que uma agente sofra uma lavagem cerebral ou seja manipulada de alguma maneira, por hipnose ou por um ataque alienígena que a faz ter um conjunto de preferências psicológicas diferentes daquele que ela teria. Em todos esses casos, que podemos chama-los de **manipulação**, Frankfurt parece estar comprometido com a visão de que essas agentes agem livremente e são moralmente responsáveis desde que o entrosamento psicológico apropriado esteja em jogo, independetemente do tipo de história que tenha gerado o entrosamento, mesmo que essa história, aparentemente, elimine a liberdade e a responsabilidade.

Isso destaca a real dificuldade que a posição de Frankfurt enfrenta à luz do Argumento a favor do Incompatibilismo de Origem. Uma vez que Frankfurt defende que, para a responsabilidade moral, a importância reside na relação entre o desejo efetivo de primeira-ordem de uma agente e sua volição de segunda-ordem, e não a origem desses respectivos desejos, o autor se compromete a pensar que, mesmo que os desejos sejam implantados por alguma força alienígena, a agente ainda poderia ser responsável por agir de acordo com esses desejos. No entanto, muitas pessoas acham duvidoso que uma agente possa ser responsável por uma ação que surge de motivos que podem ser rastreados até forças que lhe são externas. Em resposta, Frankfurt precisa mostrar que ser a origem de suas próprias motivações não é necessário para a responsabilidade moral. Se ele não for capaz de mostrar isso, então parece que o incompatibilismo de origem fica com a vantagem. A(o) incompatibilista pode simplesmente argumentar que, se um tipo de história causal alienígena que dá origem ao entrosamento frankfurtiano elimina a liberdade e a responsabilidade da agente, então por que uma história determinística não faria o mesmo?

4.3. A Perspectiva da Razão

Em *Freedom within Reason* (1990), como também em vários artigos provocativos (1980, 1987), Susan Wolf desenvolve uma teoria do livre-arbítrio e da

responsabilidade moral que destaca um entrosamento entre a ação de uma agente e o que ela (corretamente) considera valioso. Para Wolf, o livre-arbítrio diz respeito à capacidade de uma agente de agir de acordo com a Verdade e o Bem. Diferentemente de Frankfurt, as condições da teoria do entrosamento de Wolf requerem uma referência externa ao estado psicológico interno da agente (a Verdade e o Bem). A questão crucial para Wolf diz respeito ao problema de saber se uma agente é capaz de agir com base em razões morais. Logo, Wolf adota o título *A Perspectiva da Razão*. Uma versão similar, mas com diferenças importantes, foi defendida recentemente por Nelkin (2011).

Em seu esforço para fazer com que o livre-arbítrio rastreie razões morais, Wolf (1980; 1990, p. 79-81), e depois Nelkin, desenvolve uma surpreendente tese assimétrica, segundo a qual a conduta digna de mérito não requer a liberdade de agir de outro modo, mas a conduta digna de culpa exige. Colocando em termos de controle de direcionamento e controle regulador, apenas condutas culpáveis requerem o controle regulador. O controle de direcionamento é suficiente para a conduta digna de mérito. O raciocínio de Wolf é que, se uma agente atua de acordo com a Verdade e o Bem e, se de fato ela é psicologicamente determinada a fazer isso de modo que não pode deixar de agir de acordo com a Verdade e o Bem, então sua incapacidade de agir de outro modo não ameaça o tipo de liberdade que as agentes moralmente responsáveis precisam. Pois como sua liberdade poderia ser de alguma forma aprimorada simplesmente adicionando uma capacidade de agir irracionalmente? Mas a conduta culpável, argumenta Wolf, exige controle regulador, pois, se uma agente atua de maneira contrária à Verdade e ao Bem, mas é psicologicamente determinada a fazer isso de modo que não pode agir de acordo com a Verdade e o Bem, então, sendo incapaz de agir conforme a razão exige, não seria razoável culpá-la.

Como a visão assimétrica de Wolf exige o controle regulador no caso de ações culpáveis, seu compatibilismo fica vulnerável à refutação por argumentos incompatibilistas projetados para mostrar que o determinismo é incompatível com a liberdade que envolve possibilidades alternativas. Consequentemente, Wolf argumenta que determinação física não implica determinação psicológica. Assim, a autora argumenta que a determinação física é consistente com a habilidade de agir de outro modo, porque a habilidade relevante é aquela que exige apenas a falsidade do determinismo psicológico, que é uma tese que ela acredita não ter qualquer base (curiosamente, isso significa que, embora Wolf seja uma compatibilista em relação à culpabilidade e ao determinismo físico, ela é uma incompatibilista em relação à culpabilidade e ao determinismo psicológico).

E o Argumento a favor do Incompatibilismo de Origem? Na perspectiva de Wolf, se uma agente atua por razões e se as suas razões são a Verdade e o Bem (ou são suscetíveis à Verdade e ao Bem), então ela, como agente, é uma origem da conduta que carrega em si (ou é capaz de carregar em si) a marca da razão moral. Já foi dito o suficiente. Mas o que acontece com o Argumento a favor do Incompatibilismo de Origem e a premissa sobre o caráter último que parece atormentar quase todo tipo de compatibilismo: uma pessoa age de acordo com o seu livre-arbítrio apenas se for a sua origem última? Assim como a teoria do entrosamento de Frankfurt, a teoria de Wolf também é ameaçada pela ideia de que uma agente poderia ser manipulada artificialmente de uma maneira que eliminasse sua responsabilidade para satisfazer as demandas de sua teoria. E não poderia tal manipulação ser diferente da maneira como um mundo determinístico molda uma agente para ter a estrutura e a motivação psicológica que possui? A possibilidade de casos de manipulação não mostra que, sem o caráter último, uma agente não pode ser a origem adequada de sua ação? Parece que Wolf está na mesma encruzilhada de Frankfurt. Ou ela deve mostrar o que é problemático nos casos de manipulação, a fim de distinguir agentes manipuladas do tipo de entrosamento adequado exigido por sua teoria, ou ela deve encarar e aceitar que esses tipos de agentes manipuladas, pelas condições de sua teoria, agem livremente e são moralmente responsáveis por suas condutas.

4.4. Compatibilismo de Responsividade a Razões

Muitas(os) compatibilistas sugeriram que as ações livres surgem de aspectos volicionais da agência que são sensíveis a uma variedade de razões apropriadas (DENNETT, 1984a; FINGARETTE, 1972; GERT; DUGGAN, 1979; GLOVER, 1970; MACINTYRE, 1957; NEELY, 1974; NOZICK, 1981). As agentes que não respondem a considerações racionais apropriadas (como agentes compulsivas ou neuróticas) não agem livremente. Mas as agentes que **respondem** a uma variedade de considerações racionais agem livremente. Essa perspectiva foi muito bem refinada nos últimos anos por Fischer (1987, 1994) e, posteriormente, Fischer e Ravizza (1998). Para uma discussão mais avançada da perspectiva de Fischer e Ravizza, que muitas e muitos consideram o padrão de qualidade para o compatibilismo contemporâneo, consulte a seção B do complemento **Compatibilismo: o Estado da Arte**.

Uma *Teoria de Responsividade a Razões* recorre às características disposicionais da relação de uma agente com as razões relevantes em uma ação livre. Uma conduta de responsividade a razões adequadas é sensível a considerações racionais. É importante ressaltar que o ponto não é apenas que, em algumas situações contrafactuais, a agente se mostraria responsiva a razões, mas sim que sua capacidade de resposta a razões em algumas situações contrafactuais é evidência de que sua própria conduta atual, as causas que **a** originaram, ocorre também em resposta a considerações racionais (alterações precisam ser feitas para acomodar casos de ações no calor do momento, ou ações livres impulsivas).

4.4.1. Responsividade a Razões baseada no Agente

A maneira mais natural de entender uma teoria de responsividade a razões é em termos da capacidade de um **agente** de responder a razões. Para ilustrar, suponha que Frank Zappa toque banjo livremente. De acordo com uma teoria de responsividade a razões, o fato de ele tocar banjo livremente naquele momento exige que, se em ao menos alguns casos hipotéticos ele tivesse motivos para não o fazer, então ele se absteria de tocar banjo. Por exemplo, se Jimi Hendrix tivesse entrado no estúdio de gravação de Frank e pedido que ele tocasse sua guitarra, Frank gostaria de deixar Jimi feliz e, assim, alegremente deixaria seu banjo de lado e pegaria a guitarra. Parece, então, que para Frank tocar banjo livremente, Frank, **o agente**, deve ter controle regulador e não meramente controle de direcionamento sobre sua ação de tocar. A liberdade dele deve consistir parcialmente em sua habilidade de agir de acordo com alternativas.

4.4.2. Uma Tensão entre Capacidade de Resposta a Razões e os Exemplos de Frankfurt

Note que, porque os exemplos de Frankfurt desafiam a demanda incompatibilista por controle regulador, eles também desafiam as teorias de responsividade a razões **baseadas no agente** (Fischer; Ravizza, 1998, p. 34-41). Imagine que o demônio benevolente Jerry Garcia queira que Frank toque banjo no momento relevante. Jerry preferiria que Frank tocasse o banjo livremente. Porém, preocupado com a possibilidade de Frank escolher não tocar banjo, Jerry secretamente

altera as coisas para que, se for necessário, ele possa manipular Frank. Se Frank mostrar qualquer indicação de que não tocará banjo, então Jerry o manipulará de modo que Frank tocará banjo. Logo, quando ele toca banjo sem a influência da possível intervenção do outro, ele o faz livremente. Mas ele não tem controle regulador, nem parece ser responsivo a razões, em relação ao seu ato de tocar banjo. Devido à presença de Jerry, ele não pode deixar de tocar banjo, **mesmo que Jimi Hendrix peça a Frank que toque sua guitarra**.

Para aliviar a tensão entre uma teoria de responsividade a razões e exemplos de Frankfurt, Fischer argumentou que o compatibilismo de responsividade a razões pode ser moldado de tal forma que envolva somente o controle regulador. Considere o exemplo com Frank, Jimi e Jerry. Frank não tinha controle regulador sobre a sua ação de tocar banjo, dado que a presença de Jerry garantia que Frank tocaria o banjo mesmo que Jimi pedisse a Frank que ele tocasse a sua guitarra. O cenário em que Jimi pede a Frank para não tocar banjo é um cenário que Frank normalmente **consideraria** como uma razão convincente para se abster de tocar banjo. Logo, **por seus próprios motivos**, Frank **consideraria** o pedido de Jimi persuasivo. No entanto, devido à presença de Jerry, **Frank** não é responsivo a uma razão tão persuasiva. O que seria necessário para ilustrar a capacidade de resposta seria retirar Jerry do cenário. Isso já seria suficiente. Então, suponha que Frank toque banjo livremente, mesmo com Jerry passivamente a postos. Como pode ser demonstrado que a conduta de Frank foi, de alguma maneira, responsiva a razões? Como pode ser demonstrado que o que ele realmente fez foi em resposta a uma razão? Bem, *se* Jimi Hendrix tivesse pedido a Frank para não tocar banjo, mas sim a sua guitarra, *e* se a presença de Jerry fosse retirada da situação, **então** Frank responderia ao pedido de Jimi e tocaria guitarra e não banjo. Isso sugere que Frank toca banjo livremente, mesmo na situação atual na qual Jerry está passivamente a postos.

4.4.3. Uma Teoria de Responsividade a Razões baseada em Mecanismos

A ilustação da responsividade a razões em um exemplo de Frankfurt parece exigir o reconhecimento de condições contrafactuais nas quais um agente atua de outro modo em resposta a razões. Mas, em um exemplo de Frankfurt, é preciso retirar dessas condições a presença das condições de garantia (o interventor contrafactual) projetadas para garantir que o agente não aja de outro modo. Como esse passo pode ser legítimo? Como isso não é apenas um adendo arbitrário para

juntar dois temas compatibilistas que, ao contrário, parecem estar em desacordo (responsividade a razões e exemplos de Frankfurt)? Não é arbitrário, e aqui temos o porquê. Pense no que acontece no cenário atual de um exemplo de Frankfurt. À medida que as coisas se desenrolam, o demônio fica **inativo**. O agente atua por suas próprias razões. Mas agora, concentrando-se apenas **no que o agente faz** nesse cenário atual e nas razões que lhe dão uma base para fazer o que faz, considere quais características deliberativas de sua agência desempenharam o papel causal na **sequência real** de eventos que acarretam a sua ação. Portanto, concentre-se apenas no que constitui aquela gama mais estreita de características agenciais dentro do espectro mais amplo de todas as características que compunham Frank Zappa, o agente. Dado que é apenas esse espectro mais restrito que propomos identificar com a produção causal da conduta de Frank, chamemos simplesmente de **mecanismo** que produz sua ação.

Depois de localizarmos o mecanismo de ação que ocorre na sequência causal real de um exemplo de Frankfurt, podemos nos concentrar em entender as características disposicionais **dele** como um mecanismo causal. Se outras razões têm influência sobre isso, então o mecanismo seria sensível a algumas dessas razões. Isso produziria uma conduta diferente em uma variedade razoável de casos. Se esse for o caso, **então o próprio mecanismo é responsivo a razões**. Confirmar que o próprio mecanismo é responsivo a razões não apenas ilustraria que, em cenários diferentes do atual, o agente atua sobre um mecanismo sensível a razões. Também ilustraria que, no cenário de Frankfurt em que o agente **realmente** age, o que **realmente** desempenha um papel na sequência causal atual de sua ação é uma característica de **sua** agência (um mecanismo) que, **de fato**, é **em si** uma resposta a uma razão.

Fischer oferece uma análise de controle de orientação de **responsividade a razões**, **baseada em mecanismos e de sequência real**. Ele sustenta que sua análise de controle de orientação é compatível com o determinismo. De acordo com Fischer, um agente e o mecanismo de sua ação podem ser completamente determinados na sequência atual de eventos nos quais ele age. No entanto, a maneira atual pela qual o seu mecanismo responde a razões poderia ser apropriadamente sensível a razões de modo que, se razões diferentes estivessem envolvidas, o mecanismo responderia de maneira diferente e o agente que tem esse mecanismo agiria de maneira diferente da que agiu.

Recentemente, porém, Carolina Sartorio (2016) argumentou que um compatibilista da sequência real não precisa apelar à ideia de mecanismos como

meio para reivindicar suas teorias. Segundo Sartorio, podemos explicar a responsabilidade de Frank apelando para as causas reais do seu comportamento, que ela afirma serem muito mais ricas do que foi percebido anteriormente. Na proposta de Sartorio, o comportamento de Frank não é causado apenas por sua sensibilidade a razões, mas também é causado pela **ausência** de outras razões (nesse caso, por exemplo, a ausência do pedido de Jimi é em si uma causa na sequência real).

4.4.4. Avaliando o Compatibilismo de Responsividade a Razões

O compatibilismo de responsividade a razões de Fischer começa rejeitando a premissa principal do Argumento a favor do Incompatibilismo Clássico, na medida em que rejeita a tese de que liberdade do tipo exigido para a responsabilidade moral requer habilidade para agir de outro modo. Mas como a visão de Fischer se levanta contra o Argumento a favor do Incompatibilismo de Origem? O desafio aqui enfrentado por Fischer, que é colocado por Pereboom (2001) e Mele (2019), é o mesmo que o enfrentado por Frankfurt e Wolf. O incompatibilista de origem sustenta que o agente ser a origem última de sua ação é uma condição necessária para o livre-arbítrio. A tarefa do compatibilista consiste em mostrar que seu tratamento sobre a origem da conduta de um agente é suficiente para o livre-arbítrio. Mas o incompatibilista de origem vai apontar para casos de manipulação que sugerem que algumas histórias causais que acarretam estruturas psicológicas amigáveis ao compatibilista, como os mecanismos responsivos a razões, eliminam a liberdade e a responsabilidade. Se esse é o caso, então por que o determinismo é de algum modo diferente de um caso de manipulação? O ônus de mostrar como os casos de manipulação se diferenciam de uma história determinística normal parece estar com a(o) compatibilista. A única estratégia restante para a(o) compatibilista consiste simplesmente em negar que as agentes manipuladas de maneira relevante não sejam livres e moralmente responsáveis. É claro que esse problema não passou despercebido para as(os) compatibilistas de responsividade a razões. Para uma tentariva de tratar de tal assunto, confira Fischer (2004).

4.5. Compatibilismo Strawsoniano

Por fim, consideraremos as posições inspiradas pelo compatibilismo de Strawson.

Muitas filósofas contemporâneas e muitos filósofos contemporâneos propuseram temas Strawsonianos. Por exemplo, Gary Watson (1987) buscou elaborar esses temas pensando em nossas práticas sobre a responsabilidade moral e, em particular, nas atitudes moralmente reativas, juntamente com as linhas de uma teoria baseada em comunicação na qual a competência de uma agente moralmente responsável consiste, de algum modo, em ser uma interlocutora em potencial de conversas morais entre ela e a comunidade moral da qual ela faz parte. De acordo com essa perspectiva, a condição de controle para a responsabilidade moral deve acomodar a capacidade de comunicação moral, por meio de palavras e atos, com os membros da comunidade moral.

Em um desenvolvimento um pouco diferente do pensamento Strawsoniano, John Martin Fischer e Mark Ravizza defendem que uma explicação do controle de direcionamento ajuda a fornecer as condições de aplicação do conceito de responsabilidade moral, conceito que, de acordo com os autores, é Strawsoniano (1998, p. 1-27). Fischer e Ravizza formulam sua teoria Strawsoniana como um aperfeiçoamento de uma sugestão de Strawson de que a responsabilidade moral deve ser associada às reações daquelas pessoas que pertencem à comunidade moral para com os membros desta mesma comunidade. Os autores propõem que a responsabilidade moral seja desenvolvida pensando em termos das **condições de adequação** para as atitudes moralmente reativas.

Susan Wolf defende (com ressalvas significativas) a tese Strawsoniana de que o ponto de vista interpessoal (que permite o acesso às atitudes moralmente reativas) é tal que agentes livres não conseguem deixar de lado (1981). Wolf diverge em alguns pontos da maneira como Strawson defende essa visão. Mas a tese central de Wolf é Strawsoniana. Uma pessoa não consegue renunciar completamente ao ponto de vista das atitudes interpessoais e esse ponto de vista é o ponto de vista de onde nossas atitudes moralmente reativas ganham forma e força em nossa conduta. Para analisar algumas defesas relacionadas à tese de Strawson, consulte Shabo (2012) e Coates (2018).

Paul Russell (1995) também defendeu uma forma de compatibilismo Strawsoniano. Ele encontrou as características principais desse tipo de compatibilismo já nos escritos de Hume sobre o livre-arbítrio e a responsabilidade moral. Segundo Russell, podemos aprender com Hume, como fez Strawson, e entender nossas

práticas de responsabilidade moral fundamentalmente como um assunto sobre nossos sentimentos e nossas expectativas sociais como estruturadas e sustentadas por esses sentimentos. Dar atenção à nossa natureza moral, como deveríamos, dissipa qualquer suposição de que o determinismo, de alguma forma, ameaçaria as nossas concepções de liberdade e responsabilidade morais.

R. Jay Wallace (1994) oferece uma extensão da estratégia geral de Strawson em termos de normas morais de justiça para responsabilizar as pessoas de modo que se reflete em nossas práticas de responsabilidade moral (1994, p. 103-9). Dessas normas morais, e não meramente dos fatos naturalistas de que temos essas práticas, Wallace prossegue a estratégia de Strawson para esclarecer as condições exigidas para ser responsável. A posição de Wallace surgiu como uma alternativa séria aos tipos de teorias do problema do livre-arbítrio que tomam como ponto de partida teórico a natureza de pessoas ou as características dos processos relevantes para as teorias da ação que vem de ações livres.

E, finalmente, talvez a defesa recente mais detalhada do compatibilismo Strawsoniano se deve a David Shoemaker (2017). Shoemaker argumenta por uma posição dependente de respostas sobre a responsabilidade moral defendendo, primeiro, uma posição dependente de respostas sobre o divertimento. Ele argumenta, então, que as normas do humor são bastante próximas àquelas de emoções que acarretam a responsabilidade, como a raiva. E, a partir disso, ele constrói sua teoria sobre a responsabilidade moral que é dependentede-respostas.

Referências

ADAMS, R. M. Involuntary Sins. **Philosophical Review**, v. 94, p. 3-31, 1985.

ARPALY, N. **Merit, Meaning, and Human Bondage**. Princeton: Princeton University Press, 2006.

ARPALY, N. **Unprincipled Virtue**. New York: Oxford, 2003.

AYER, A. J. Freedom and Necessity. *In:* WATSON (ed.). **His Philosophical Essays.** New York: St. Martin's Press, 1982, p. 3-20.

AYER, A. J. Free-Will and Rationality. *In:* VAN STRAATEN, Z. (ed.). **Philosophical Subjects**: Essays Presented to P. F. Strawson. Oxford: Clarendon, 1980.

BEATY, M. (ed.). **Christian Theism and the Problems of Philosophy**. Notre Dame: University of Notre Dame Press, 1990.

BENNETT, J. Accountability. *In:* VAN STRAATEN, Z. (ed.). **Philosophical Subjects**: Essays Presented to P. F. Strawson. Oxford: Clarendon, 1980.

BEROFSKY, B. **Nature's Challenge to Free Will**. Oxford: Oxford University Press, 2012.

BEROFSKY, B. Ifs, Cans, and Free Will: The Issues. *In:* KANE, R. (ed.). **The Oxford Handbook of Free Will**. Oxford: Oxford University Press, 2002.

BEROFSKY, B. **Liberation from Self**. Cambridge: Cambridge University Press, 1995.

BEROFSKY, B. **Freedom from Necessity.** London: Routledge & Kegan Paul, 1987.

BEROFSKY, B. (ed.). **Free Will and Determinism**. New York: Harper & Row, 1966.

BJORNNSON, G.; PERSSON, K. The Explanatory Component of Moral Responsibility. **Nous**, v. 46, p. 326-354, 2012.

BOK, H. **Freedom and Responsibility**. Princeton: Princeton University Press: 1998.

BRATMAN, M. **Structures of Agency.** New York: Oxford. 2007

BRATMAN, M. Three Theories of Self-governance. **Philosophical Topics,** v. 32, p. 21-46, 2004.

BRATMAN, M. A Desire of One's Own. **Journal of Philosophy**, v. 100, n. 5, p. 221-242, 2003.

BRATMAN, M. **Faces of Intention:** Selected Essays on Intention and Agency. Cambridge: Cambridge University Press, 1999.

BRATMAN, M. Responsibility and Planning. **Journal of Philosophy**, v. 1, n. 1, p. 27–43, 1997.

BUSS, S.; OVERTON, L. **Contours of Agency**: Essays on Themes from Harry Frankfurt. Cambridge: MIT Press, 2002.

CAMPBELL, J. A Compatibilist Theory of Alternative Possibilities. **Philosophical Studies**, v. 88, p. 319-330, 1997.

CHISHOLM, R. **Human Freedom and the Self.** The Lindley Lectures. Kansas: Department of Philosophy, University of Kansas, 1964.

CLARKE, R. Dispositions, Abilities to Act, and Free Will: The New Dispositionalism. **Mind**, v.118, n. 470, p. 323-51, 2009.

CLARKE, R. **Incompatibilist (Nondeterministic) Theories of Free Will**. *In:* ZALTA, N. (ed.). The Stanford Encyclopedia of Philosophy. Fall Edition. Stanford, CA: The Metaphysics Research Lab, 2000. Disponível em: https://plato.stanford.edu/archives/fall2000/entries/incompatibilism-theories/. Acesso em: 23 ago. 2021.

CLARKE, R. Toward a Credible Agent-Causal Account of Free Will. **Noûs**, v. 27, p.191-203, 1993.

COATES, D. J. Hard Incompatibilism and the Participant Attitude. **Canadian Journal of Philosophy**, v. 49, p. 208-229, 2018.

DENNETT, D. **Freedom Evolves**. London: Penguin Books, 2003.

DENNETT, D. **Elbow Room:** The Varieties of Free Will Worth Wanting. Cambridge: MIT Press, 1984a.

DENNETT, D. I Could Not Have Done Otherwise - So What? **The Journal of Philosophy**, v. LXXXI, n. 10, p. 553-567, 1984b.

DENNETT, D. **Brainstorms**: Philosophical Essays on Mind and Psychology. Cambridge: MIT Press, 1981a.

DENNETT, D. Conditions of Personhood. 1981b. *In:* DENNETT, D. **Brainstorms**: Philosophical Essays on Mind and Psychology. Cambridge: MIT Press, 1981a, p.267-286.

DENNETT, D. On Giving Libertarians What They Say They Want. 1981c. *In:* DENNETT, D. **Brainstorms**: Philosophical Essays on Mind and Psychology. Cambridge: MIT Press, 1981a, p. 286-299.

DENNETT, D. Mechanism and Responsibility. *In:* LEHRER, K. (ed.). **Freedom and Determinism**. New York: Random House, 1966.

DORR, C. Against Counterfactual Miracles. **Philosophical Review**, v. 125, p. 241-286, 2016.

ESHLEMAN, A. S. Being is not Believing: Fischer and Ravizza on Taking Responsibility. **Australasian Journal of Philosophy**, v.79, p. 479-490, 2001.

FARA, M. Masked Abilities and Compatibilism. **Mind**, v. 117, n. 468, p. 843-865, 2008.

FINGARETTE, H. **The Meaning of Criminal Insanity**. Berkeley: University of California Press, 1972.

FISCHER, J. M. Semicompatibilism and Its Rivals. **Journal of Ethics**, v. 16, p. 117-143, 2012.

FISCHER, J. M. Freedom, Foreknowledge, and Frankfurt: A Reply to Vihvelin. **Canadian Journal of Philosophy**, v. 38, n. 3, p. 327-342, 2008.

FISCHER, J. M. **My Way**. New York: Oxford University Press, 2006.

FISCHER, J. M. Reasons Reactivity & Incompatibilist Intuitions. **Philosophical Explorations**, v. 8, n. 2, p. 131-43, 2005.

FISCHER, J. M. Responsibility and Manipulation. **Journal of Ethics**, v. 8, n. 2, p. 145-77, 2004.

FISCHER, J. M. Recent Work on Moral Responsibility. **Ethics**, v. 110, p. 93-139, 1999.

FISCHER, J. M. **The Metaphysics of Free Will**. Oxford: Blackwell Publishers, 1994.

FISCHER, J. M. Responsiveness and Moral Responsibility. *In:* SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 1987, p. 81-106.

FISCHER, J. M. (ed.). **Moral Responsibility**. Ithaca: Cornell University Press, 1986.

FISCHER, J. M. Power Necessity. **Philosophical Topics**, v. 14, p. 77-91, 1986.

FISCHER, J. M. Incompatibilism. **Philosophical Studies**, v. 43, p. 127-37, 1983.

FISCHER, J. M. Responsibility and Control. **Journal of Philosophy**, v. 79, p. 24-40, 1982.

FISCHER, J. M.; KANE, R.; PEREBOOM, D.; VARGAS, M. **Four Views on Free Will**. Malden: Blackwell Publishers, 2007.

FISCHER, J. M.; RAVIZZA, M. **Responsibility and Control**: An Essay on Moral Responsibility. Cambridge: Cambridge University Press, 1998.

FISCHER, J. M.; RAVIZZA, M. (ed). **Perspectives on Moral Responsibility**. Ithaca: Cornell University Press, 1993.

FRANKFURT, H. *In:* BUSS, S.; OVERTON, L. **Contours of Agency**: Essays on Themes from Harry Frankfurt. Cambridge: MIT Press, 2002, p. 27-31.

FRANKFURT, H. **Necessity, Volition, and Love**. Cambridge: Cambridge University Press, 1999.

FRANKFURT, H. **Autonomy, Necessity, and Love**. Cambridge: Cambridge University Press, 1994.

FRANKFURT, H. **The Importance of What We Care About**. Cambridge: Cambridge University Press, 1988.

FRANKFURT, H. Identification and Wholeheartedness. *In:* SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 1987.

FRANKFURT, H. Freedom of the Will and the Concept of a Person. **Journal of Philosophy**, v. 68, p. 5-20, 1971.

FRANKFURT, H. Alternate Possibilities and Moral Responsibility. **Journal of Philosophy**, v. 66, p. 829-839, 1969.

FULDA, H. F.; HORSTMANN, R. P. (ed.). **Vernunftbegriffe in der Moderne:** Stuttgart Hegel-Kongress 1993. Stuttgart: Klett-Cotta, 1994.

GERT, B.; DUGGAN, T. Free Will as the Ability to Will. **Noûs**, v. 13, p. 197-217, 1979.

GINET, C. Libertarianism. LOUX, M.; ZIMMERMAN, D. (ed.). **Oxford Handbook of Metaphysics**. Oxford: Oxford University Press, 2003.

GINET, C. In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing. **Philosophical Perspectives**, v. 10, p. 403-417, 1996.

GINET, C. **On Action**. Cambridge: Cambridge University Press, 1990.

GINET, C. In Defense of Incompatibilism. **Philosophical Studies**, v. 44, p. 391-400, 1983.

GINET, C. The Conditional Analysis of Freedom. *In:* VAN INWAGEN, P. (ed.). **Time and Cause**. Dordrecht: D. Reidel,1980.

GINET, C. Might We Have No Choice? *In:* LEHRER, K. (ed.). **Freedom and Determinism**. New York: Random House, 1966, p. 87-104.

GLOVER, J. **Responsibility**. New York: Humanities Press, 1970.

GRAHAM, P. A. A Defense of Local Miracle Compatibilism. **Philosophical Studies**, v.140, p. 65-82, 2008.

HAJI, I. **Incompatibilism's Allure**. Peterborough. Ontario: Broadview Press, 2009.

HAJI, I. Compatibilist Views of Freedom and Responsibility. *In:* KANE, R. (ed.). **The Oxford Handbook of Free Will**. Oxford: Oxford University Press, 2002.

HAJI, I. **Moral Appraisability**. New York: Oxford University Press, 1998.

HAJI, I.; CUYPERS, S. **Moral Responsibility, Authenticity, and Normative Education**. New York: Routledge Press, 2008.

HOBART, R. E. Free Will as Involving Indeterminism and Inconceivable Without It. **Mind**, v.43, p. 1-27, 1934.

HOBBES, T. Leviathan. *In:* FLATMAN, R. E.; JOHNSTON, D. (ed.). **Leviathan**. New York: W.W. Norton & Co, 1997.

HONDERICH, T. **A Theory of Determinism**. Oxford: Clarendon Press, 1988.

HONDERICH, T. (ed.). **Essays on Freedom and Action**. London: Routledge & Kegan Paul, 1973.

HOWARD-SNYDER, D.; JORDAN, J. (ed.). **Faith, Freedom, and Rationality**. Lanham: Rowman and Littlefield, 1996.

HUME, D. **An Enquiry Concerning Human Understanding**. Oxford: Clarendon Press, 1975.

HUME, D. **A Treatise of Human Nature**. Oxford: Clarendon Press, 1978.

HUNT, D. P. Moral Responsibility and Unavoidable Action. **Philosophical Studies**, v. 97, p. 195-227, 2000.

KANE, R. (ed.). **The Oxford Handbook of Free Will**. Oxford: Oxford University Press, 2002.

KANE, R. (ed.). **The Significance of Free Will.** Oxford: Oxford University Press, 1996.

KAPITAN, T. A Master Argument for Incompatibilism? *In:* KANE, R. (ed.). **The Oxford Handbook of Free Will**. Oxford: Oxford University Press, 2002, p. 127-57.

KAPITAN, T. Autonomy and Manipulated Freedom. **Philosophical Perspectives**, v.14, p. 81-104, 2000.

KENNETT, J. **Agency and Responsibility.** New York: Oxford University Press, 2001.

LAMB, J. On a Proof of Incompatibilism. **Philosophical Review**, v. 86, p. 20-35, 1977.

LEHRER, K. (ed.). **Freedom and Determinism**. New York: Random House, 1966.LENMAN, J. Compatibilism and Contractualism: The Possibility of Moral Responsibility. **Ethics**, v. 117, p. 7-31, 2006.

LEVY, N.; MCKENNA, M. Recent Work on Moral Responsibility. **Philosophy Compass**, v. 43, p. 96-133, 2008.

LEWIS, D. Are We Free to Break the Laws? **Theoria**, v. 47, p. 113-21, 1981.

LOUX, M.; ZIMMERMAN, D. (ed.). **Oxford Handbook of Metaphysics**. Oxford: Oxford University Press, 2003.

MACINTYRE, A. Determinism. **Mind**, v. 66, p. 28-41, 1957.

MCKAY, T.; JOHNSON, D. A Reconsideration of an Argument against Incompatibilism. **Philosophical Topics**, v. 24, p. 113-22, 1996.

MCKENNA, M. Reasons-Responsiveness, Agents, and Mechanisms. *In:* SHOEMAKER, D. (ed.). **Oxford Studies in Agency and Responsibility**. v. 1. Oxford: Oxford University Press, 2013, p. 151-183.

MCKENNA, M. Ultimacy & Sweet Jane. *In:* TRAKAKIS, N.; COHEN, D. (ed.). **Essays on Free Will and Moral Responsibility**. UK: Cambridge Scholars Publishing, 2008, p. 186-208.

MCKENNA, M. Reasons Reactivity & Incompatibilist Intuitions. **Philosophical Explorations**, v. 8, n. 2, p. 131-143, 2005.

MCKENNA, M. Responsibility and Globally Manipulated Agents. **Philosophical Topics**, v. 32, p. 169-82, 2004.

MCKENNA, M. Robustness, Control, and the Demand for Morally Significant Alternatives. *In:* WIDERKER, D.; MCKENNA, M. (ed.). **Moral Responsibility and Alternative Possibilities**. Aldershot: Ashgate Press, 2003.

MCKENNA, M. Review of Hilary Bok's, Freedom & Responsibility. **Ethics**, v. 113, n. 1, p. 144-145, 2002.

MCKENNA, M. Critical notice of Ishtiyaque Haji's, Moral Appraisability. **Philosophy and Phenomenological Research**, v. 63, p. 711-715, 2001a.

MCKENNA, M. John Martin Fischer and Mark Ravizza's Responsibility & Control Review. **Journal of Philosophy**, v. XCVIII, n. 2, p. 93-100, 2001b.

MCKENNA, M. Assessing Reasons-Responsive Compatibilism. **International Journal of Philosophical Studies**, v. 8, n. 1, p. 89-114, 2000.

MCKENNA, M. The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism. **Journal of Ethics**, v. 2, p. 123-142, 1998.

MCKENNA, M. Alternative Possibilities and the Failure of the Counterexample Strategy. **Journal of Social Philosophy**, v. 28, p. 71-85, 1997.

MELE, A. **Free Will and Luck**. New York: Oxford University Press, 2006.

MELE, A. Reactive Attitudes, Reactivity, and Omissions. **Philosophy and Phenomenological Research**, v. 61, p. 447-452, 2000.

MELE, A. **Autonomous Agents**. New York: Oxford University Press, 1995.

MELE, A.; ROBB, D. Rescuing Frankfurt-Style Cases. **Philosophical Review**, v. 107, p. 97-112, 1998.

NAHMIAS, E.; MORRIS, S.; NADELHOFFER, T.; TURNER, J. Is Incompatibilism Intuitive? **Philosophy and Phenomenological Research,** v. 73, p. 28-53, 2006.

NAYLOR, M. B. Frankfurt on the Principle of Alternate Possibilities. **Philosophical Studies**, v. 46, p. 249-58, 1984.

NEELY, W. Freedom and Desire. **Philosophical Review**, v. 83, p. 32-54, 1974.

NELKIN, D. K. **Making Sense of Freedom and Moral Responsibility**. Oxford: Oxford University Press, 2011.

NICHOLS, S. Folk Intuitions on Free Will. **Journal of Cognition and Culture**, v. 6, n. 1-2, p. 57-86, 2006.

NOZICK, R. **Philosophical Explanations**. Cambridge: Harvard University Press, 1981.

O'CONNOR, T. Free Wil. *In:* ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy**. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2002. Disponível em: https://plato.stanford.edu/archives/spr2002/entries/freewill/. Acesso em: 24 ago. 2021.

O'CONNOR, T. **Persons and Causes**. New York: Oxford University Press, 2000.

O'CONNOR, T. On the Transfer of Necessity. **Noûs**, v. 27, p. 204-218, 1993.

OTSUKA, M. Incompatibilism and the Avoidability of Blame. **Ethics**, v. 108, p. 685-701, 1998.

PENDERGRAFT, G. The Explanatory Power of Local Miracle Compatibilism. **Philosophical Studies**, v. 156, p. 249-266, 2011.

PEREBOOM, D. **Living Without Free Will**. Cambridge: Cambridge University Press, 2001.

PEREBOOM, D. Determinism al Dente. **Noûs**, v. 29, p. 21-45, 1995.

PERRY, J. Compatibilist Options. *In:* CAMPBELL, J.; O'ROURKE, M.; SHIER, D. (ed.). **Freedom and Determinism**. Cambridge: MIT Press: 2004, p. 231-254.

PETTIT, P. **A Theory of Freedom**: From the Psychology to the Politics of Agency. New York: Oxford University Press, 2001.

PETTIT, P.; SMITH, M. Freedom in Belief and Desire. **Journal of Philosophy**, v. 93, n. 9, p. 429-449, 1996.

ROSKIES, A.; NICHOLS, S. Bringing Moral Responsibility Down to Earth. **Journal of Philosophy**, v. 105, n. 7, p. 371-388, 2008.

ROWE, W. **Thomas Reid on Freedom and Morality**. Ithaca: Cornell University Press, 1991.

RUSSELL, P. Critical Notice of John Martin Fischer and Mark Ravizza Responsibility and Control: A Theory of Moral Responsibility. **Canadian Journal of Philosophy**, v. 32, p. 587-606, 2002a.

RUSSELL, P. Pessimists, Pollyannas, and the New Compatibilism. 2002b. *In:* KANE, R. (ed.). **The Oxford Handbook of Free Will**. Oxford: Oxford University Press, 2002, p. 229-256.

RUSSELL, P. **Freedom and Moral Sentiment**. New York: Oxford University Press, 1995.

RUSSELL, P. Strawson's Way of Naturalizing Responsibility. **Ethics**, v. 102, p. 287-302, 1992.

SARTORIO, C. **Causation and Free Will**. Oxford: Oxford Universit Press, 2016.

SAUNDERS, J. T. The Temptation of Powerlessness. **American Philosophical Quarterly**, v. 5, p. 100-108, 1968.

SCANLON, T. M. **Moral Dimensions**: Permissibility, Meaning, Blame. Cambridge: Belknap Harvard Press, 2008.

SCANLON, T. M. **What We Owe to Each Other**. Cambridge: Harvard University Press, 1998.

SCHLICK, M. When is a Man Responsible? *In:* BEROFSKY, B. (ed.). **Free Will and Determinism**. New York: Harper & Row, 1966, p. 54-62.

SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 1987.

SHABO, S. Where Love and Resentment Meet: Strawson's Interpersonal Defense of Compatibilism. **The Philosophical Review**, v. 121, p. 95-124, 2012.

SHOEMAKER, David. Response-Dependent Responsibility; or, a Funny Thing Happened on the Way to Blame. **The Philosophical Review**, v. 126, p. 481-527, 2017.

SLOTE, M. Understanding Free Will. **Journal of Philosophy**, v. 77, p. 136-51, 1980.

SLOTE, M. Review of Peter van Inwagen's An Essay on Free Will. **Journal of Philosophy**, v. 82, p. 327-30, 1985.

SLOTE, M. Selective Necessity and the Free-Will Problem. **Journal of Philosophy**, v. 79, p. 5-24, 1982.

SMART, J. J. C. Free Will, Praise, and Blame. **Mind**, v. 70, p. 291-306, 1963.

SMITH, A. Responsibility for Attitudes: Activity and Passivity in Mental Life. **Ethics**, v. 115, p. 236-71, 2005.

SMITH, M. Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion. *In:* STROUD, S.; TAPPOLET, C. (ed.). **Weakness of Will and Practical Irrationality**. New York: Oxford University Press, 2003, p. 17-38.

STRAWSON, G. **Freedom and Belief**. Oxford: Clarendon Press, 1986.

STRAWSON, P. F. Freedom and Resentment. **Proceedings of the British Academy**, v. 48, p. 187-211, 1962.

STROUD, S.; TAPPOLET, C. (ed.). **Weakness of Will and Practical Irrationality**. New York: Oxford University Press, 2003.

STUMP, E. Control and Causal Determinism. *In:* BUSS, S.; OVERTON, L. **Contours of Agency**: Essays on Themes from Harry. Frankfurt: Cambridge; MIT Press, 2002, p. 33-60.

STUMP, E. Libertarian Freedom and the Principle of Alternative Possibilities. 1996a. *In:* HOWARD-SNYDER, D.; JORDAN, J. (ed.). **Faith, Freedom, and Rationality**. Lanham: Rowman and Littlefield, 1996, p. 73-88.

STUMP, E. Persons: Identification and Freedom. **Philosophical Topics**, v. 24, p.183-214, 1996b.

STUMP, E. Intellect, Will, and the Principle of Alternate Possibilities. *In:* FISCHER, J. M.; RAVIZZA, M. (ed). **Perspectives on Moral Responsibility**. Ithaca: Cornell University Press, 1993, p. 237-262.

TAYLOR, R. **Metaphysics**. Englewood Cliffs: Prentice Hall, 1974.

TODD, P.; Tognazzini, N. A Problem for Guidance Control. **Philosophical Quarterly**, v. 58, n. 233, p. 685-692, 2008.

TRAKAKIS, N.; COHEN, D. (ed.). **Essays on Free Will and Moral Responsibility**. UK: Cambridge Scholars Publishing, 2008.

VAN INWAGEN, P. Free Will Remains a Mystery? *In:* KANE, R. (ed.). **The Oxford Handbook of Free Will.** New York: Oxford University Press, 2002, p. 158-177.

VAN INWAGEN, P. **An Essay on Free Will**. Oxford: Clarendon Press, 1983.

VAN INWAGEN, P. (ed.). **Time and Cause**. Dordrecht: D. Reidel,1980.

VAN INWAGEN, P. Ability and Responsibility. **Philosophical Review**, v. 87, p. 201-224, 1978.

VAN INWAGEN, P. The Incompatibility of Free Will and Determinism. **Philosophical Studies**, v. 27, p. 185-99, 1975.

VAN STRAATEN, Zak (ed.). **Philosophical Subjects**: Essays Presented to P. F. Strawson. Oxford: Clarendon, 1980.

VARGAS, M. Moral Influence, Moral Responsibility. *In:* TRAKAKIS, N.; COHEN, D. (ed.). **Essays on Free Will and Moral Responsibility**. UK: Cambridge Scholars Publishing, 2008, p. 90-122.

VARGAS, M. The Revisionist's Guide to Responsibility. **Philosophical Studies**, v. 125, p. 399-429, 2005.

VELLEMAN, J. D. **The Possibility of Practical Reason**. Oxford: Oxford University Press, 2000.

VELLEMAN, J. D. What Happens When Someone Acts? **Mind**, v. 101, p. 462-481, 1992.

VIHVELIN, K. **Causes, Laws, & Free Will**. New York: Oxford University Press, 2013.

VIHVELIN, K. Foreknowledge, Frankfurt, and the Ability to Do Otherwise: A Reply to Fischer. **Canadian Journal of Philosophy**, v. 38, p. 3, p. 343-72, 2008.

VIHVELIN, K. Free Will Demystified: A Dispositional Account. **Philosophical Topics**, v. 32, p. 427-50, 2004.

VIHVELIN, K. Freedom, Foreknowledge, and the Principle of Alternate Possibilities. **Canadian Journal of Philosophy**, v. 30 p. 1-24, 2000.

WALLACE, R. J. **Responsibility and the Moral Sentiments**. Cambridge: Harvard University Press, 1994.

WATSON, G. **Agency and Answerability**. New York: Oxford University Press, 2004.

WATSON, G. Reason and Responsibility. **Ethics**, v. 111, p. 374-394, 2001.

WATSON, G. Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. *In:* SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 1987, p. 256-286.

WATSON, G. (ed.). **Free Will**. New York: Oxford University Press,1982.

WATSON, G. Free Agency. **Journal of Philosophy**, v. 72, p. 205-220, 1975.

WIDERKER, D. On an Argument for Incompatibilism. **Analysis**, v. 47, p. 37-41, 1987.

WIDERKER, D. Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. **Philosophical Review**, v.104, p. 247-261, 1995.

WIDERKER, D.; MCKENNA, M. (ed.). **Moral Responsibility and Alternative Possibilities**. Aldershot: Ashgate Press, 2003.

WIGGINS, D. Towards a Reasonable Libertarianism. *In:* HONDERICH, T. (ed.). **Essays on Freedom and Action**. London: Routledge & Kegan Paul, 1973, p. 31-62.

WOLF, S. **Freedom within Reason.** Oxford: Oxford University Press, 1990.

WOLF, S. Sanity and the Metaphysics of Responsibility. *In:* SCHOEMAN, F. (ed.). **Responsibility, Character, and the Emotions**: New Essays in Moral Psychology. Cambridge: Cambridge University Press, 1987, p. 45-64.

WOLF, S. The Importance of Free Will. FISCHER, J. M.; RAVIZZA, M. (ed). **Perspectives on Moral Responsibility**. Ithaca: Cornell University Press, 1993, p. 386-405.

WOLF, S. Asymmetrical Freedom. **Journal of Philosophy**, v. 77, p. 157-66, 1980.

WYMA, K. Moral Responsibility and Leeway for Action. **American Philosophical Quarterly**, v. 34, p. 57-70, 1997.

ZIMMERMAN, D. Reasons-Responsiveness and the Ownership of Agency: Fischer and Ravizza's Historicist Theory of Responsibility. **Journal of Ethics**, v. 6, p. 199-234, 2002.

ZIMMERMAN, M. J. **An Essay on Moral Responsibility**. Totowa, NJ: Rowman and Littlefield, 1988.

(IV) Teorias Contrafactuais da Causação*

Autores: Peter Menzies & Helen Beebee
Tradutores: Renato Semaniuc Valvassori & Ulisses Franceschi Eliano
Revisores Científicos: Karenina do Nascimento Rodrigues, Kherian Gracher & Pedro Merlussi

A ideia básica por trás das teorias contrafactuais de causação é que o significado de afirmações causais pode ser explicado em termos de condicionais contrafactuais na forma "Se **A** não tivesse ocorrido, **C** não teria ocorrido". Boa parte das análises contrafactuais propostas focaram em afirmações na forma "evento **c** causou o evento **e**", descrevendo causação "singular", "espécime" ["*token*"] ou "atual". Tais análises se tornaram populares desde o desenvolvimento da semântica de mundos possíveis para contrafactuais na década de 1970. A análise contrafactual da causação mais conhecida é a teoria de David Lewis (1973b). Contudo, uma intensa discussão ao longo de quarenta anos colocou em dúvida a adequação de

---

* MENZIES, P.; BEEBEE, H. Counterfactual Theories of Causation. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2020. Disponível em: https://plato.stanford.edu/archives/win2020/entries/causation-counterfactual/. Acesso em: 01 ago. 2021.

The following is the translation of the entry on Scientific Method, 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/win2020/entries/causation-counterfactual/>. 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/archives/win2020/entries/causation-counterfactual/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry.

qualquer análise simples de causação singular em termos de contrafactuais. Nos últimos anos, tem ocorrido uma proliferação de diferentes maneiras de se refinar a ideia central; a modelagem da "equação estrutural" ou "modelamento causal" é atualmente a maneira mais popular de se resgatar a relação entre causação e contrafactuais.

1. A Análise de Lewis para Contrafactuais de 1973.

A ideia norteadora por trás da análise contrafactual da causação é a de que, como coloca David Lewis (LEWIS, 1973b, p. 161), "[...] pensamos causa como sendo algo que faz uma diferença, e a diferença que ela faz deve ser a diferença daquilo que teria ocorrido sem ela. Se ela tivesse se ausentado, seus efeitos – alguns deles, ao menos, mas usualmente todos – teriam igualmente estado ausentes."

A primeira definição explícita de causação em termos de contrafactuais foi surpreendentemente oferecida por Hume, quando este escreveu: "Nós podemos definir causa como sendo **um objeto seguido por outro e, onde todos os objetos, similares ao primeiro, são seguidos por objetos similares ao segundo**. Ou, em outras palavras, **onde, se o primeiro objeto não tivesse existido, o segundo também não teria existido.**" (1748, Seção VII). É difícil entender como Hume poderia ter confundido a primeira, definição por regularidade, com a segunda e bem distinta, definição contrafactual (*vide* BUCKLE, 2004, p. 212-213, para uma breve discussão).

De qualquer forma, Hume nunca explorou a abordagem contrafactual alternativa para causação. Nisso, como em muitas outras coisas, ele foi seguido por gerações de filósofas e filósofos empiristas. O principal obstáculo da explicação de causação em termos de contrafactuais na concepção das e dos empiristas era a obscuridade das contrafactuais em si, devida principalmente à referência a possibilidades não atuais. O potencial verdadeiro da abordagem contrafactual à causação não se tornou claro até que houve uma melhor compreensão das contrafactuais, através do desenvolvimento da semântica de mundos possíveis, no início da década de 1970.

A mais conhecida e minuciosamente elaborada teoria contrafactual da causação é aquela proposta por David Lewis (1973b). O autor a refinou e a expandiu em artigos subsequentes reunidos em (1986a). Em resposta às dúvidas quanto ao tratamento dado pela sua teoria da preempção, ele propôs posteriormente uma revisão radical de sua própria pesquisa (2000, 2004a). Nesta seção, limitaremos

nossa atenção à teoria original de 1973, adiando as alterações posteriores que ele propôs para as considerações abaixo.

1.1. Contrafactuais e Dependência Causal

Assim como a maioria das teorias contemporâneas de contrafactuais, a teoria de Lewis emprega uma semântica de mundos possíveis para contrafactuais. Tal semântica estabelece condições de verdade para contrafactuais em termos de relações de similaridade entre mundos possíveis. Lewis defende o célebre realismo sobre mundos possíveis, de acordo com o qual mundos possíveis não atuais são entidades concretas em pé de igualdade com o mundo atual (LEWIS, 1986e). No entanto, boa parte das filósofas contemporâneas e dos filósofos contemporâneos gostaria de implantar a estrutura prolífica de mundos possíveis, ao mesmo tempo em que se distanciam do realismo completo acerca deles[4] .

A noção central da semântica de mundos possíveis para contrafactuais é a de relação de **similaridade comparativa** entre mundos (LEWIS, 1973a). Diz-se de um mundo que é **mais próximo da realidade** do que outro se o primeiro se assemelha ao mundo atual mais do que o segundo. Em termos dessa relação de similaridade, a condição de verdade para a contrafactual "Se **A** fosse (ou tivesse sido) o caso, **C** seria (ou teria sido) o caso" é estabelecida como se segue:

> (1) "Se **A** fosse o caso, **C** seria o caso" é verdadeira no mundo atual se, e somente se, ou (i) não há **mundos-A** possíveis; ou (ii) algum **mundo-A** onde **C** se sustenta é mais próximo do mundo atual do que qualquer **mundo-A** onde **C** não se sustenta.

Devemos ignorar o primeiro caso no qual a contrafactual é verdadeira por vacuidade. A ideia fundamental dessa análise é que a contrafactual "Se **A** fosse o caso, **C** seria o caso" é verdadeira apenas quando o distanciamento da realidade

---

[4] Vide: MENZEL, C. Possible Worlds. *In*: ZALTA, N. E. (ed.). T**he Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: Stanford, CA: The Metaphysics Research Lab, 2017. Disponível em: https://plato.stanford.edu/archives/win2017/entries/possible-worlds/. Acesso em: 23 ago. 2021.

para tornar o antecedente verdadeiro junto do consequente é menor do que o distanciamento para tornar o antecedente verdadeiro sem o consequente.

Em termos de contrafactuais, Lewis define uma noção de dependência causal entre eventos que tem um papel importante em sua teoria de causação (1973b).

> (2) Sejam **c** e *e* dois eventos possíveis distintos, tal que **e depende causalmente** de **c** se, e somente se, na ocorrência de **c**, o evento **e** ocorreria; e, caso **c** não ocorresse, **e** não ocorreria.

Essa condição estabelece que a ocorrência de **e** depende da de **c**. Onde **c** e **e** são eventos que ocorrem atualmente, a condição de verdade pode ser de certa forma simplificada. Nesse caso, segue-se da segunda condição formal da relação de similaridade comparativa que a contrafactual "Se *c* ocorresse, *e* ocorreria" é automaticamente verdadeira: essa condição formal implica que uma contrafactual com antecedente e consequente verdadeiros é ela própria também verdadeira. Consequentemente, a condição de verdade para dependência causal torna-se:

> (3) Sejam **c** e **e** eventos distintos atuais, tal que **e** é **depende causalmente** de **c** se, e somente se, caso **c** não ocorresse, o evento **e** não ocorreria.

Há três coisas importantes a se observar sobre a definição de dependência causal. Primeiramente, ela considera os *relata* básicos de dependência causal como sendo eventos. A própria Teoria de Eventos de Lewis (1986b) constrói eventos como sendo classes de regiões espaço-temporais possíveis. Contudo, diferentes concepções de eventos são compatíveis com a definição básica, como as de Kim (1973a). Para uma posição alternativa, em boa parte lewisiana, sobre eventos, consulte McDonnell (2016) e Kaiserman (2017). Na realidade, até mesmo parece possível formular a definição básica em termos de fatos ao invés de eventos, como sugere Mellor (1995; 2004).

Em segundo lugar, a definição requer que os eventos causalmente dependentes sejam **distintos** uns dos outros. Distinção significa que os eventos não são idênticos, nem são parte um do outro ou implicam um ao outro. Essa qualificação é importante no caso de se descartar dependências não-causais espúrias (*vide* KIM, 1973b; LEWIS, 1986b). Desse modo, você não teria escrito "Larry" se não tivesse escrito "rr", e você não teria dito em alto e bom tom "Olá" se você não tivesse dito "Olá". Nenhuma dessas dependências conta como dependência causal, dado que os pares de eventos não são distintos uns dos outros no sentido requerido.

Convencido da necessidade de dar espaço à sua análise de causação por (e da) ausência – como quando um jardineiro causa a morte de suas plantas ao deixar de regá-las – Lewis, posteriormente, adiciona à sua visão a ideia de que dependência causal é uma questão de dependência contrafactual entre eventos ou suas ausências (LEWIS, 2000, p. §X; 2004b). Devemos ignorar essa complicação no que se segue. Para alguma discussão sobre causação pela ausência ver Schaffer (2000b), Beebee (2004b), McGrath (2005), Livengood e Machery (2007) e Dowe (2009).

Em terceiro lugar, as contrafactuais que são empregadas nessa análise devem ser entendidas a partir daquilo que Lewis chama de interpretação padrão. Há diversas maneiras de se interpretar contrafactuais; e algumas interpretações geram dependências não-causais espúrias entre eventos. Por exemplo, suponha que os eventos **c** e **e** são efeitos de uma causa comum **d**. É tentador pensar que deve haver uma dependência causal entre *c* e *e* ao se envolver com o seguinte raciocínio contrafactual: se *c* não tivesse ocorrido, então *d* não teria ocorrido; e se **d** não tivesse ocorrido, *e* não teria ocorrido. Mas Lewis diz que a primeira contrafactual, a qual ele chama de contrafactual **retroativa**[5] , não deve ser usada na avaliação de dependência causal. As contrafactuais que de fato devem ser usadas são as contrafactuais irretroativas que tipicamente mantém o passado fixo até o momento (ou até pouco antes do momento) no qual o antecedente da contrafactual supostamente sustenta. Assim, se *c* não tivesse ocorrido, **d** – que na realidade ocorreu antes de **c** – teria ocorrido de qualquer maneira; portanto, na interpretação padrão, onde as contrafactuais retroativas são falsas, a inferência de que *e* depende causalmente de *c* é bloqueada.

1.2. A Assimetria Temporal da Dependência Causal

O que constitui a direção da relação causal? Por que essa direção é tipicamente alinhada com a direção temporal que vai do passado ao futuro? Respondendo a essas perguntas, Lewis (1979) argumenta que a direção da causação é a direção da dependência causal; e é tipicamente verdadeiro que eventos dependem causalmente de eventos anteriores, mas não de eventos posteriores. Lewis enfatiza

---

[5] N.T.: Optamos por traduzir a expressão *backtracking contrafactual* por "contrafactual retroativa".

a contingência desse último fato por considerar causação retroativa ou temporalmente reversa como uma possibilidade conceitual que não pode ser descartada *a priori*. Consequentemente, ele dispensa qualquer análise de contrafactuais que assumisse a assimetria temporal por meio de um postulado conceitual.

A explicação de Lewis para a assimetria temporal da dependência contrafactual provém da combinação de sua análise da relação de similaridade, junto da (alegada) "assimetria da sobredeterminação", uma característica contingente do mundo. De acordo com essa análise, há uma série de aspectos da similaridade que devem ser levados em conta em avaliações de contrafactuais irretroativas: similaridade, no que diz respeito às leis da natureza, e, também, similaridade concernente a fatos particulares. Os mundos são mais similares ao mundo atual quanto menores forem os milagres e violações das leis da natureza atuais neles contidos. Novamente, quanto maior for a região espaço-temporal que faça a correspondência perfeita entre os fatos particulares dos mundos com o mundo atual, mais similares são esses mundos do mundo atual. Se as leis da natureza do mundo atual são determinísticas, essas regras se confrontarão na hora de acessar quais mundos são mais similares ao mundo atual, pois um mundo que torna o antecedente da contrafactual verdadeira deve diferir do mundo atual ou permitir a violação das leis atuais (um "milagre de divergência"), ou diferir do mundo atual no que diz respeito a fatos particulares. A análise de Lewis permite, nesses casos, balancear esses dois aspectos da similaridade. Isso implica que mundos, com uma extensa região de correspondência perfeita dos fatos particulares, podem ser considerados bastante similares ao mundo atual, concedido que a correspondência com os fatos particulares do mundo atual é alcançada com o custo de um pequeno milagre local, mas não ao custo de um grande milagre de divergência.

Tomada por si só, essa explicação não contém uma assimetria temporal em si mesma. Isso se dá apenas quando tal explicação é combinada com a assimetria da sobredeterminação: o (alegado) fato de que os efeitos são raramente sobredeterminados por suas causas, mas causas são frequentemente sobredeterminadas por seus efeitos. Tomando um exemplo de Elga (2000): suponha que Gretta quebre um ovo às 8h (evento **c**), estoure-o na frigideira e o coma em seu café da manhã. O que teria acontecido caso **c** não tivesse ocorrido? A resposta certa (resposta 1) é que o ovo não teria sido frito e Gretta não o teria comido, mas não (resposta 2) que ela ainda assim teria fritado e comido o ovo, e que esses eventos teriam de alguma forma se dado apesar de ela não ter quebrado o ovo num primeiro momento. A questão é: como a análise de Lewis da relação de similaridade oferece a resposta

1, e não a resposta 2? Particularmente, considere mundos onde há uma correspondência perfeita entre fatos particulares até um instante antes das 8h e, então, após a ocorrência de um milagre, nos momentos que se seguem deixa de existir uma correspondência perfeita entre os fatos particulares desses mundos. Vamos chamar o mundo possível mais próximo, tal qual descrito, de Mundo 1. Agora, considere mundos onde não há uma correspondência perfeita entre fatos particulares antes das 8h (e, em particular, Gretta não quebra o ovo), e há um milagre logo **após** as 8h. Então, passa a existir uma correspondência perfeita entre fatos particulares. Chamemos o mundo mais próximo, tal qual descrito, de Mundo 2 (intuitivamente, no primeiro caso, mantemos o passado fixo, inserimos um milagre logo antes das 8h de modo que **c** não ocorra, e o futuro se desdobra em seguida de acordo com as leis (atuais). No segundo caso, mantemos o futuro fixo, inserimos um milagre logo após as 8h, de modo que **c** não ocorra, e o passado se desenrola de acordo com as leis (atuais)). Por que seria o Mundo 1 mais próximo da atualidade que o Mundo 2?

A resposta de Lewis a esta questão surge do fato de que **c** deixa muitos rastros: às 8h02min, por exemplo, há um ovo sendo feito na panela, a casca quebrada se encontra na lixeira, há vestígios de ovo cru nos dedos de Gretta. Esta tem a memória de tê-lo acabado de quebrar, e assim por diante. Assim, no Mundo 2, Gretta deixa de quebrar o ovo, mas, em seguida, parece lembrar de tê-lo quebrado, há um ovo na panela, a casca do ovo está na lixeira, e tudo o mais. Portanto, o Mundo 2, dado que possui todos esses eventos mesmo sem que o ovo tenha sido quebrado inicialmente, deve conter não apenas um milagre, mas uma série deles: para cada um desses efeitos, um milagre é requerido como explicação. O Mundo 1, por outro lado, requer apenas um pequeno milagre que faz com que Gretta não quebre o ovo. Consequentemente, o Mundo 2 contém um milagre "grande e diverso", enquanto o Mundo 1 possui apenas um pequeno milagre, por conseguinte, o Mundo 1 é mais próximo da atualidade que o Mundo 2. Sendo assim, a análise de Lewis fornece o resultado correto de que se Gretta não tivesse quebrado o ovo, ela não o teria comido.

O resultado no caso de Gretta é generalizado na medida em que causas são sobredeterminadas por seus efeitos, mas efeitos não são sobredeterminados por suas causas. Sobredeterminação de efeitos pelas causas, apesar de realmente ocorrerem, como quando a vítima leva, simultaneamente, tiros de vários assassinos, mas são relativamente raros e, mesmo quando ocorrem, os efeitos são sobredeterminados apenas por um grupo de eventos. Por outro lado, os rastros deixados estão sempre presentes, e, ou ao menos Lewis assim pensa, a extensão da sobredeterminação, em qualquer caso, é maior do que em casos de sobredeterminação da causa para o

efeito. Contudo, ambas são características contingentes do mundo atual, ou ao menos é o que Lewis defende (*vide* § 2.1).

Em geral, portanto, a análise simétrica de similaridade e a assimetria *de facto* da sobredeterminação implicam, juntas, que mundos que acomodam mudanças contrafactuais, preservando o passado e permitindo milagres de divergência, são mais similares ao mundo atual do que mundos que acomodam tais mudanças, permitindo milagres de convergência que preservam o futuro atual. Esse fato, por sua vez, implica que, onde a assimetria de sobredeterminação se sustenta, o presente depende contrafactualmente do passado, mas não do futuro.

1.3. Transitividade e Preempção

Como observa Lewis (1973b), dependência causal entre eventos atuais é suficiente para causação, mas não é necessária: é possível ter causação sem haver dependência causal. Um caso padrão de "preempção" ilustrará esse ponto. Suponha que dois atiradores conspirem para o assassinato de um odiado ditador, concordando que um dos dois atirará no ditador em uma ocasião pública. Agindo um ao lado do outro, assassinos **A** e **B** encontram um local de boa visibilidade e, quando aparece o ditador, ambos miram (eventos **a** e **b**, respectivamente). *A* pressiona o gatilho e atira, acertando seu alvo, enquanto **B** desiste de atirar ao ver **A** pressionando o gatilho. Aqui as ações do assassino **A** (como a de mirar) são causas da morte do ditador, enquanto as ações de **B** (tal qual sua ação de mirar) são apenas causas potenciais preemptivas (Lewis distingue tais casos de preempção dos casos de sobredeterminação causal, nos quais dois processos culminam no efeito, sem que quaisquer dos processos se antecipem ao outro. Lewis acredita que esses casos não são testes adequados para uma teoria da causação, dado que não provocam julgamentos claros). O problema levantado por esse exemplo de preempção é que ambas as ações estão no mesmo nível do ponto de vista da dependência causal: não tivessem **A** nem **B** agido, o ditador não teria morrido, e se apenas um tivesse agido, o ditador teria morrido.

De modo a superar esse problema, Lewis estende a dependência causal para uma relação transitiva tomando seu ancestral. Ele define uma **cadeia causal** como uma sequência finita de eventos atuais **c**, **d**, **e**…onde **d** depende causalmente de **c**, **e** depende de **d**, e assim por diante no decorrer da sequência. A causação é finalmente definida nesses termos:

(4) **c** é uma causa de **e** se, e somente se, existe uma cadeia causal atual levando de **c** até **e**.

Dada a definição de causação em termos de cadeias causais, Lewis é capaz de distinguir entre causas atuais preemptivas (como **a**) das causas potenciais preemptivas (como **b**). Há uma cadeia causal indo de *a* até a morte do ditador, mas não há uma cadeia análoga indo de **b** até a morte do ditador. Tome, por exemplo, como um evento intermediário ocorrendo entre **a** e a morte do ditador, a bala da arma de **A** acelerando no ar no meio da trajetória. A aceleração da bala depende causalmente de **a**, dado que essa bala particular não estaria no meio da trajetória caso **A** não tivesse mirado; e a morte do ditador depende causalmente da aceleração da bala, dado que no meio da trajetória da bala, **B** se absteve de atirar, de modo que o ditador não teria morrido na ausência da aceleração da bala (lembre-se de que não nos é permitido "retroceder": não é verdade que se a bala não tivesse estado no meio da trajetória, **A** não teria mirado e, consequentemente, não é verdade que se a bala não tivesse estado no meio da trajetória, **B** teria, afinal, disparado). Consequentemente, temos uma cadeia causal e, portanto, causação. Mas nenhum evento intermediário correspondente, entre **b** e a morte do ditador, pode ser encontrado, consequentemente, **b** não conta como uma das causas de sua morte.

A definição de causação de Lewis também apresenta como resultado que causação é uma relação transitiva: para todo **c** que causa **d**, e **d** que causa **e**, será verdadeiro que **c** causa **e**. A transitividade da causação se encaixa com ao menos uma parte de nossas práticas explanatórias. Por exemplo, historiadoras e historiadores, querendo explicar algum evento histórico significativo, traçarão uma explicação que remonta a uma série de conexões causais, concluindo que o evento no início da cadeia causal é responsável pela explicação do evento histórico que está sendo estudado. Não obstante, como veremos depois, algumas autoras e alguns autores alegaram que causação não é de fato transitiva.

1.4. Causação Incerta[6]

Até agora, consideramos como a teoria contrafactual da causação funciona quando se assume o determinismo. Mas, e quanto à causação nos casos em que

---

[6] N.T.: Optamos por traduzir *chancy causation* por "causação incerta". Apenas na ocorrência do termo *probabilistic*, optamos por traduzir por "probabilística(o)".

o determinismo falha? Lewis (1986c) argumenta que causação incerta é uma possibilidade conceitual que deve ser acomodada por uma teoria da causação. De fato, a física contemporânea nos informa que o mundo atual está repleto de processos probabilísticos que possuem caráter causal. Para dar um exemplo conhecido (LEWIS, 1986c): suponha que você, maliciosamente, conecta uma bomba a uma fonte radioativa e a um contador Geiger, de modo que a bomba explode se o contador registrar certo número de cliques num período de dez minutos. Se acontecer de o contador registrar o número requerido de cliques e a bomba explodir, seu ato será a causa da explosão, ainda que não haja qualquer conexão determinística entre seu ato e a explosão da bomba: seria consistente com o passado atual e as leis da natureza que o contador Geiger não pudesse ter registrado o número de cliques suficiente.

Em princípio, a análise contrafactual da causação funciona bem para lidar com causação incerta, dado que dependência contrafactual não requer que a causa seja suficiente, nas circunstâncias, para o efeito, a análise contrafactual apenas requer que a causa tenha sido necessária, nas circunstâncias, para o efeito. O problema que se coloca ao abandonar a hipótese do determinismo, contudo, é que o indeterminismo preponderante mina a plausibilidade da ideia de que, deixando de lado preempção e sobredeterminação, efeitos em geral dependem contrafactualmente de suas causas. No caso do contador Geiger mencionado acima, por exemplo, suponha que a chance de que a bomba exploda pode ser alterada por meio de um indicador (uma configuração baixa significa que o contador Geiger precisa registrar muitos cliques para que a bomba dispare em dez minutos, tornando a explosão, consequentemente, muito improvável). O indicador está em uma configuração baixa, assim, eu aumento as chances de a bomba explodir ao aumentá-la. Meu ato foi a causa da explosão, mas não é verdade que, se eu não tivesse feito isso, a bomba não teria explodido, visto que a explosão apenas teria sido bastante improvável de ocorrer.

De modo a acomodar a causação incerta, Lewis (1986c) define uma noção mais geral de dependência causal, em termos de contrafactuais incertos. Esses contrafactuais são da forma “Se A fosse o caso, Pr(C) seria x”, em que a contrafactual é uma contrafactual subjuntiva comum, interpretada de acordo com a semântica acima, e o operador **Pr** é um operador de probabilidade com escopo reduzido, limitado ao consequente da contrafactual. Lewis interpreta as probabilidades envolvidas como chances únicas indexadas temporalmente (*vide* LEWIS, 1980).

A noção mais geral de dependência causal pode ser lida como se segue:

(5) Sejam **c** e **e** eventos atuais distintos, tal que *e* depende causalmente de **c** se, e somente se, não tivesse **c** ocorrido, a chance do evento **e** ocorrer seria muito inferior à chance atual.

Essa definição abrange casos de causação determinística nos quais a chance do efeito com a causa é 1, e a chance do efeito sem a causa é 0. Mas ela também permite casos de causação probabilística irredutível, onde essas chances podem tomar valores não extremos, como no exemplo do contador Geiger apresentado acima. É semelhante à noção central de relevância probabilística, usada em teorias probabilísticas de causação-tipo, excetuando-se o fato de que ela se utiliza de contrafactuais incertas em vez de probabilidades condicionais. Veja a discussão em Lewis (1986c) para as vantagens da abordagem contrafactual sobre a abordagem probabilística.

O restante da teoria de causação incerta segue o perfil da teoria determinística da causação: novamente, tem-se causação quando possuímos uma ou mais etapas de dependência causal.

2. Problemas para a Teoria Contrafactual de Lewis

Nesta seção, exploraremos as principais dificuldades da teoria de Lewis, que passaram a ser discutidas nos últimos quarenta e cinco anos.

2.1. Assimetria temporal

Houve uma série de discussões criticando a explicação de Lewis para a assimetria temporal da causação (HORWICH, 1987, cap. 10; HAUSMAN, 1998, cap. 6; PRICE, 1996, cap. 6). Uma importante crítica diz respeito à "assimetria de milagres", que é um ponto central da abordagem de Lewis à assimetria temporal da causação: um milagre que satisfaz o antecedente de uma contrafactual sobre fatos particulares no tempo, por ter um mundo possível divergente do mundo atual imediatamente antes do tempo t é, afirma Lewis, menor e menos diverso que um milagre que satisfaz o antecedente da mesma contrafactual e faz um mundo possível convergir com o mundo atual após o tempo t. Adam Elga (2000) argumentou que a assimetria de milagres não se sustenta em muitos casos.

O argumento de Elga prossegue por meio do exemplo de Gretta quebrando um ovo, descrito anteriormente, e a ideia básica é de que realmente o Mundo 2, acima, mais próximo onde os fatos particulares do passado, até pouco depois das 8h, não correspondem aos fatos particulares atuais (e, em particular, Gretta não quebra o ovo), há um milagre logo após as 8h e, depois disso, o Mundo 2 se desenvolve de acordo com as leis atuais, passando a corresponder perfeitamente no que diz respeito aos fatos particulares atuais. Pensando na direção que vai do futuro ao passado, considere o que ocorre no mundo atual de um ponto de vista do tempo reverso: Gretta tira o ovo de seu prato e o coloca na frigideira quente, a frigideira esfria e ovo se torna cru pulando, em seguida, para dentro da casca que se fecha ordenadamente ao seu redor. Agora (para nos levar ao Mundo 2) insira um pequeno milagre, às 8h05min, por exemplo, quando o ovo já está feito na frigideira, alterando apenas a posição de algumas poucas moléculas, de modo que o que ocorre (novamente, procedendo do futuro para o passado) é que (de acordo com as leis, uma vez que o milagre ocorreu) o ovo apenas fica na frigideira, esfriando e transferindo o calor a ela no processo e, então, gradualmente apodrece como normalmente ocorre com os ovos (excetuando-se que ovos normalmente apodrecem na direção passado-para-o-futuro). A ideia é que, enquanto o Mundo 2 (visto do passado para o futuro) parece excessivamente estranho – afinal de contas, ele envolve Gretta devorando o que foi um horrível ovo estragado que de alguma forma foi parar em sua frigideira e bizarramente desapodreceu – nada disso está em desacordo com as leis. É apenas, dadas as leis da termodinâmica, extremamente improvável.

Mas como isso nos ajuda com o problema original relacionado ao Mundo 2, a saber, o fato de que todos os rastros da Gretta quebrando o ovo atualmente, tais como ela se lembrando de tê-lo quebrado, a presença da casca de ovo vazia e quebrada na lixeira, e assim por diante, teriam de ser, de alguma forma, provocados no Mundo 2 por meio de milagres adicionais, de modo a preservar a correspondência perfeita entre fatos particulares do Mundo 2 com os do mundo atual após às 8h05min? A resposta curta seria que não nos ajuda. Os "rastros", de fato, estão no Mundo 2 e, é claro que, normalmente, esperaríamos que tais "rastros" apontassem conclusivamente para Gretta tendo recentemente quebrado um ovo. Mas eles não implicam de acordo com as leis que ela o fez. Novamente, olhando para o Mundo 2 por meio de uma perspectiva de tempo reverso, tomamos o mundo tal qual ele é a partir de 8h05min, com seus "rastros" e demais características, e fazemos as leis retrocederem (a despeito do pequeno milagre que faz o ovo apodrecer lentamente na panela em vez de fazê-lo pular para sua casca). O que se "segue" (ainda retrocedendo no

tempo) é uma incógnita e, seja o que for, sem dúvida parecerá bizarro quando visto na direção usual do passado para o futuro. Seja como for, o Mundo 2 é um mundo com um único milagre pequeno às 8h05min, com uma correspondência perfeita entre seus fatos particulares e os do mundo atual no período que se segue, o que o torna tão próximo do mundo real quanto o Mundo 1, onde há um pequeno milagre antes de Gretta quebrar o ovo. Então, Elga argumenta, não há assimetria de dependência contrafactual, como Lewis define.

Várias autoras e autores chegaram à conclusão de que uma abordagem viável da assimetria da dependência contrafactual provém não da assimetria dos milagres, como afirma Lewis, mas da assimetria termodinâmica: isto é, do fato de que a entropia aumenta conforme se avança em direção ao futuro. Em particular, David Albert (2000) propõe que precisamos assumir a "Hipótese do Passado": a suposição de que o universo começou em uma condição de entropia extremamente baixa. Para maiores discussões sobre a Hipótese do Passado e sua relação com a assimetria da dependência contrafactual, confira Frisch (2005, 2007), Loewer (2007), Price e Weslake (2009) e Kutach (2002, 2013).

2.2. Transitividade

Como vimos, Lewis elabora a transitividade para a causação, definindo causalidade em termos de cadeias de dependência causal. No entanto, foram apresentados vários supostos contraexemplos que põem em dúvida a transitividade. A esse respeito, Lewis (2004a) apresenta uma breve classificação desses contraexemplos. Eis dois desses contraexemplos.

Primeiro, temos um exemplo não publicado, mas muito discutido graças a Ned Hall. Um trilheiro[7] está caminhando por uma trilha que atravessa uma montanha, quando uma pedra lá do alto se desprende e rola a encosta. O trilheiro percebe a pedra e dela se desvia na hora certa. A pedra que rola causa o desvio do trilheiro e isso, por sua vez, causa a continuidade de sua caminhada (essa segunda conexão causal envolve **prevenção dupla**: o desvio previne a colisão entre a pedra e o trilheiro que, caso tivesse ocorrido, teria evitado a continuidade

[7] N.T.: Utilizaremos aqui o termo "trilheiro" para *hiker*, isto é, aquele que pratica *hicking*, esporte de caminhadas, essencialmente realizado em ambientes naturais, como trilhas e estradas.

da caminhada do trilheiro). Contudo, a pedra que rola é o tipo de coisa que normalmente evitaria a continuidade da caminhada do trilheiro, então parece contraintuitivo dizer que ela causou a caminhada.

No segundo contraexemplo, oferecido por Douglas Ehring (1987), Jones coloca alguns sais de potássio em fogo quente. Como os compostos de potássio produzem uma chama roxa quando aquecidos, a chama se torna arroxeada, embora todas as demais coisas permaneçam iguais. A chama roxa, então, incendeia algum material inflamável nas proximidades. Aqui, julgamos que colocar os sais de potássio no fogo causou a chama roxa que, por sua vez, provocou a combustão do material inflamável. Mas parece implausível assumir que o ato de colocar os sais de potássio no fogo tenha causado a combustão do material inflamável.

Uma série de respostas foi elaborada a esses contraexemplos. L. A. Paul (2004) oferece uma resposta ao segundo exemplo, que envolve conceber os *relata* da causação como **aspectos** de eventos: ela argumenta que há uma incompatibilidade entre o aspecto do evento que é o efeito da primeira conexão causal (a chama sendo de cor roxa), e o aspecto do evento que é a causa da segunda conexão causal (a chama tocando o material inflamável). Assim, embora seja verdade que a chama roxa não causou a combustão, não há, afinal de contas, uma falha de transitividade. Maslen (2004) resolve o problema recorrendo a uma abordagem contrastivista da causalidade (*vide* § 4): a situação contrastante no efeito final da primeira afirmação causal não coincide com a situação de contraste na causa final da segunda afirmação causal. Portanto, a primeira afirmação causal deve ser interpretada como dizendo que o fato de Jones ter colocado sais de potássio no fogo, em vez de não o ter feito, fez com que **a chama ficasse roxa em vez de amarela**; mas a segunda afirmação causal deve ser interpretada como dizendo que **o fato de o fogo roxo ter surgido ao invés de não ter surgido** causou o material inflamável ter entrado em processo de combustão, em vez de não ter entrado. Onde existe uma incompatibilidade desse tipo, não temos um contraexemplo genuíno à transitividade.

O primeiro exemplo não pode ser tratado da mesma maneira. Algumas defensoras e alguns defensores da transitividade responderam que nossas intuições sobre a intransitividade da causação nesses exemplos são enganosas. Lewis (2004a), por exemplo, aponta que os contraexemplos à transitividade envolvem tipicamente uma estrutura na qual um evento tipo-**c** geralmente impede um evento tipo-**e**, mas, no caso em particular, o evento-**c** causa atualmente outro evento que contém a ameaça e causa o evento-**e**. Se misturarmos perguntas sobre que coisas geralmente levam a outras, com perguntas sobre o que causou o que nesse caso

em particular, ele diz, podemos pensar que é razoável negar que **c** causa **e**. Mas, se mantivermos o foco estritamente no caso em particular, devemos insistir que *c* realmente causa *e*.

O debate sobre a transitividade da causação não é facilmente resolvido, em parte porque está ligado à questão de qual a melhor forma de uma teoria contrafactual lidar com exemplos de preempção. Como vimos, a teoria contrafactual de Lewis se baseia na transitividade da causação para lidar com casos de preempção. Se esses casos pudessem ser tratados de alguma outra maneira, isso tiraria da teoria parte da pressão teórica que nela se encontra, permitindo que ela reconhecesse os alegados contraexemplos à transitividade, sem sucumbir às dificuldades colocadas pela preempção. Para mais informações sobre esse assunto, consulte Hitchcock (2001). Para uma extensa discussão sobre as questões relacionadas à transitividade, consulte Paul e Hall (2013, cap. 5).

2.3. Preempção

Como vimos, Lewis emprega sua estratégia de definir causalidade em termos de cadeias de dependência causal não apenas para tornar a causação transitiva, mas também para lidar com exemplos de preempção. No entanto, existem exemplos de preempção com os quais essa estratégia não consegue lidar satisfatoriamente. As dificuldades relacionadas à preempção provaram serem os maiores pesadelos à teoria de Lewis (Paul e Hall, 2013: Cap. 3 contém uma extensa discussão dos problemas colocados pela preempção e outros tipos de causalidade redundante para teorias contrafactuais).

Em seu (1986c: 200), Lewis distingue casos de preempção precoce e tardia[8] . Nos exemplos de preempção precoce, o processo em execução, a partir da alternativa preemptiva, é interrompido antes que o processo principal em execução, a partir da causa preemptiva, seja concluído. O exemplo dos dois assassinos, dado acima, é um exemplo desse tipo. A teoria da causalidade em termos de cadeias de dependência causal pode lidar com esse tipo de exemplo. Por outro lado, casos de preempção tardia são aqueles em que o processo em execução, a partir da causa

[8] N.T.: Optamos por traduzir "*early and late preemption*" por "preempção tardia e precoce".

preemptiva, é interrompido pelo processo principal — o qual é executado até a conclusão, produzindo o efeito — antes que a causa potencial preemptiva tenha a oportunidade de fazê-lo. A seguir, é apresentado um exemplo de preempção tardia, elaborado por Hall (2004).

Billy e Suzy jogam pedras em uma garrafa de vidro. Suzy joga primeiro, de modo que sua pedra chega primeiro e quebra o vidro; A pedra de Billy atravessa a porção de ar onde a garrafa esteve momentos antes. Se Suzy não tivesse arremessado, o arremesso de Billy teria quebrado a garrafa. No entanto, o arremesso de Suzy causou o despedaçar da garrafa, enquanto o arremesso de Billy é apenas uma causa potencial preemptiva. Esse é um caso de preempção tardia, porque o processo alternativo (lançamento de Billy) é interrompido pelo processo principal (lançamento de Suzy), que está em execução até ser concluído.

A teoria de Lewis não consegue explicar o entendimento de que o arremesso de Suzy causou o despedaçar da garrafa. Isso porque não há dependência causal entre o arremesso de Suzy e a quebra da garrafa, uma vez que, mesmo no caso em que Suzy não tivesse jogado sua pedra, a garrafa teria se quebrado devido ao arremesso de Billy. Tampouco existe uma cadeia de dependência gradual que causa o efeito, dado que não há um evento intermediário, entre o arremesso de Suzy e a quebra da garrafa, que os conecte a uma cadência de dependências. Tome, por exemplo, a pedra de Suzy no meio de sua trajetória. Esse evento depende do arremesso inicial de Suzy, mas o problema é que o despedaçar da garrafa não depende dele, posto que sem o arremesso de Suzy a garrafa ainda teria se estilhaçado por conta do arremesso de Billy.

Certamente, a quebra da garrafa, que teria ocorrido sem o arremesso de Suzy, teria sido diferente da quebra da garrafa que, realmente, ocorreu com seu arremesso. Para começar, a quebra da garrafa teria ocorrido mais tarde. Essa observação sugere que uma solução para o problema da preempção tardia pode ser insistir que os eventos envolvidos sejam interpretados como eventos frágeis. Consequentemente, será verdadeiro, e não falso que, se Suzy não tivesse jogado sua pedra, o despedaçar atual da garrafa, tomado como um evento frágil com um tempo e um modo essenciais de ocorrência, não teria ocorrido. O próprio Lewis não endossa essa resposta pelo fato de que uma política uniforme de interpretar os eventos como frágeis contrariaria nossas práticas usuais e geraria muitas dependências causais espúrias. Por exemplo, suponha que um veneno mate sua vítima mais lenta e dolorosamente quando tomado com o estômago cheio. Então, o ato da vítima de jantar antes de beber o veneno contará como causa de sua própria morte, já que

o horário e o modo da morte dependem do ato de comer o jantar (para a discussão das limitações dessa resposta, consulte Lewis, 1986c, 2000).

A solução para o problema da preempção tardia, que Lewis apoia cautelosamente em 1986c, apela à noção de quase-dependência. Considere um caso que se assemelha ao caso de Billy e Suzy jogando pedras em uma garrafa. Suzy joga uma pedra (*c*) e quebra a garrafa (*e*) exatamente da mesma maneira que ela faz no caso original. Mas, neste caso, Billy e sua pedra estão totalmente ausentes. No caso original, *e* é causado por, mas não depende contrafactualmente de *c*, enquanto no segundo caso *e* é causado por e depende contrafactualmente de *c*. Mas o caráter intrínseco do processo que leva de *c* até *e* é o mesmo nos dois casos. Assim, diz Lewis, no caso original (contando também com o arremesso de Billy), *e* quase-depende de *c*. Portanto, "poderíamos redefinir uma cadeia causal como uma sequência de dois ou mais eventos, com dependência ou quase-dependência em cada uma das etapas. E, como sempre, um evento é causa de outro se, e somente se, houver uma cadeia causal de um para o outro" (1986c, 206). (Uma ideia relacionada é adotada em Menzies, 1996, 1999). Observe que, apesar dessa definição proposta de cadeia causal, a solução de quase-dependência não exige transitividade da mesma maneira que a solução anterior de Lewis, para o problema da preempção precoce, mas exigia: com causas potenciais alternativas seguramente fora do caminho, em todos os casos de preempção (precoce e tardia), o efeito deve quase-depender diretamente de sua causa.

A insatisfação de Lewis com suas próprias tentativas de lidar com o problema da preempção tardia, bem como a incapacidade de sua teoria lidar com a "preempção preponderante"[9] (Schaffer 2000a), levou ao desenvolvimento de sua teoria de 2000. Outro problema relacionado à preempção, que surge para a causação incerta — que a teoria de 2000 não aborda —, é discutido no §5.4 abaixo.

3. A teoria de Lewis de 2000

Em uma tentativa de lidar com os vários problemas enfrentados por sua teoria de 1973, Lewis desenvolveu uma nova versão da teoria contrafactual que ele apresentou pela primeira vez em suas *Whitehead Lectures* na Universidade de

---

[9] N.T.: Optamos por traduzir "*trumping preemption*" por "preempção preponderante".

Harvard em março de 1999 (uma versão abreviada das palestras apareceu em 2000. As palestras completas foram publicadas em seu 2004a).

As contrafactuais desempenham um papel central na nova teoria, assim como na antiga. Mas as contrafactuais empregadas na nova teoria não estabelecem simplesmente a ocorrência de um evento como dependente da *ocorrência* de outro evento. As contrafactuais declaram que um evento, para ocorrer, depende da *ocorrência* ou não de outro evento e de *como* e *quando* esse outro evento ocorre. Uma ideia-chave na formulação dessas contrafactuais é a de uma *alteração* de um evento. Esse é um evento, atualizado ou não-atualizado, que ocorre em um momento ou de uma maneira ligeiramente diferente do evento em questão. Uma alteração é, por definição, um evento bastante frágil que não poderia ocorrer em um horário ou de uma maneira diferente, sem ser um evento diferente. Lewis pretende que a terminologia seja neutra quanto à questão de saber se uma alteração de um evento é uma versão do mesmo evento ou um evento numericamente distinto.

A noção central da nova teoria é a de influência:

> (6) Sejam *c* e *e* são eventos distintos, *c influencia* o evento *e* se, e somente se, existe um coleção substancial *c1*, *c2*, ... de diferentes alterações não-muito-distantes de *c* (incluindo a alteração atual de *c*) e existe uma coleção *e1*, *e2*, ... de alterações do evento *e*, as quais ao menos algumas delas se diferem, de modo que se *c1* tivesse ocorrido, *e1* teria ocorrido, e se *c2* tivesse ocorrido, *e2* teria ocorrido, e assim por diante.

Onde um evento influencia outro, há um padrão de dependência contrafactual de *se*, *quando* e *como* em consequência de *se*, *quando* e *como*.[10] Do mesmo modo que antes, a causalidade é definida como uma relação primitiva:

> *(7) c* causa *e* se, e somente se, há uma cadeia de influências graduais que levam de *c* até *e*.

Um dos pontos que Lewis avança em defesa dessa nova teoria é que ela é capaz de lidar com casos de preempção tardia e precoce (a teoria é restrita à

---

[10] Ênfases em itálico adicionadas na tradução (N. do T.).

causa determinística e, portanto, não trata do exemplo de preempção probabilística descrito abaixo em §5.4). Reconsidere, por exemplo, o exemplo de preempção tardia envolvendo Billy e Suzy jogando pedras em uma garrafa. A teoria deve explicar por que o arremesso de Suzy, e não o de Billy, é a causa do despedaçar da garrafa. Se tomarmos uma alteração, na qual o arremesso de Suzy é ligeiramente diferente (a pedra é mais leve, ou ela a lança mais cedo), mantendo fixo o arremesso de Billy, descobrimos que o despedaçar também é diferente. Mas, se tomarmos alterações semelhantes no arremesso de Billy, mantendo o arremesso de Suzy fixo, descobriremos que a quebra da garrafa permanece inalterada.

Outro ponto a favor da nova teoria é que ela lida com casos de preempção "preponderante", descritos pela primeira vez por Jonathan Schaffer (2000a). Lewis dá um exemplo que envolve um major e um sargento que estão gritando ordens aos soldados. O major e o sargento simultaneamente gritam 'Avancem!'; os soldados ouvem a ambos e avançam. Como os soldados obedecem ao oficial superior, eles avançam porque o major ordena, não porque o sargento o faz. Portanto, o comando do major suplanta, por preempção ou por preponderância, o do sargento. Outras teorias têm dificuldade com casos de preempção preponderantes, incluindo — ou ao menos Lewis assim o crê — sua própria tentativa de resolver o problema da preempção tardia apelando à quase-dependência (2000, 184-5). O caso de preponderância é aquele em que a cadeia causal, que vai do grito do sargento ao avanço dos soldados, é concluída — ou pelo menos, Lewis pensa, é epistemicamente possível que o seja — assim como a cadeia que se inicia com os gritos do major também o é. Portanto, ela é uma duplicata intrínseca do caso comparável em que o sargento grita, mas o major não; consequentemente, o avanço dos soldados quase-depende do grito do sargento, que é o resultado errado. Lewis argumenta que sua nova teoria lida facilmente com os casos de preponderância. Alterando o comando do major ao mesmo tempo em que se mantém fixo o comando do sargento, a resposta do soldado seria alterada de forma correspondente. Por outro lado, alterar o comando do sargento, mantendo fixo o comando do major, não produziria qualquer diferença.

Há, no entanto, alguma razão de ceticismo sobre se a nova teoria lida com os exemplos de preempção tardia e de preponderância de forma completamente satisfatória. No exemplo da preempção tardia, o arremesso de Billy tem algum grau de influência na quebra da garrafa. Pois, se Billy tivesse jogado sua pedra mais cedo (de modo que precedesse o arremesso de Suzy) e de uma maneira diferente, a garrafa teria quebrado mais cedo e de uma maneira diferente. Da mesma forma,

o comando do sargento tem algum grau de influência no avanço dos soldados, pois, se o sargento tivesse gritado antes do major com um comando diferente, os soldados teriam obedecido à ordem dele. Em resposta a esses pontos, Lewis deve dizer que essas alterações dos eventos são muito distantes para serem consideradas relevantes. Mas é preciso alguma forma de se medir a distância nas alterações, posto parecer que alterações semelhantes às do arremesso de Suzy e do comando do major são relevantes para terem influência causal.

Argumentou-se também que a nova teoria gera um grande número de casos espúrios de causalidade (Collins, 2000; Kvart, 2001). A teoria implica que qualquer evento que influencie a outro, em certa medida, conta como uma de suas causas. Mas o senso comum é mais judicioso ao analisar causas. Para dar um exemplo de Jonathan Bennett (1987): chuvas em dezembro atrasam incêndios florestais; se não houvesse chuvas de dezembro, a floresta pegaria fogo em janeiro e não em fevereiro. A chuva influencia o fogo no que diz respeito ao tempo, localização, rapidez e assim por diante. Mas o senso comum nega que a chuva tenha sido a causa *do incêndio*, embora permita que seja a causa do *atraso do incêndio*. Da mesma forma, no exemplo da vítima de veneno discutido acima, a ingestão de veneno pela vítima com o estômago cheio influencia o tempo e o modo de sua morte (tornando-a lenta e dolorosa), mas o senso comum se recusa a aceitar seu ato de jantar como uma causa de *sua morte*, embora possa encará-lo como causa de *ser sua morte lenta e dolorosa*. Com todo respeito à Lewis, o senso comum não toma qualquer coisa que afeta o tempo e o modo de um evento como sendo uma causa *simpliciter* do evento.

4. Contextualismo *versus* Invariantismo

Uma pergunta que tem recebido crescente atenção nos últimos anos é se a causação é uma relação "invariável" ou se, em vez disso, a verdade de uma determinada afirmação causal varia de acordo com o contexto no qual se está discutindo (observe que "invariável" é frequentemente usado para descrever uma relação causal que é estável em uma ampla gama de circunstâncias diferentes; esse não é o significado de "invariável" que está sendo utilizado aqui). Há muitas evidências de que os julgamentos causais das pessoas são sensíveis a fatores contextuais (Hilton & Slugoski, 1986; Cheng e Novick, 1991; Knobe & Fraser, 2008; Hitchcock & Knobe, 2009; Clarke *et al.*, 2015; Kominsky *et al.*, 2015; Icard *et al.*,

2017); no entanto, em princípio, pode-se manter a linha invariantista e insistir que o que varia com o contexto não é a verdade, mas apenas assertibilidade.

Considere um caso de problema padrão: o jardineiro e a rainha igualmente deixam de regar minhas flores, enquanto estou de férias, e sua morte subsequente (*e*) depende contrafactual e igualmente de suas omissões. Mas muitos dos julgamentos das pessoas apontam apenas a omissão do jardineiro — e não a da rainha — como sendo uma causa genuína de suas mortes (Beebee, 2004b; McGrath, 2005; Livengood & Machery, 2007). Podemos acomodar esse julgamento alegando que é falso que a omissão da rainha tenha sido uma causa de *e*, concluindo que precisamos de alguma restrição adicional à causação, além da dependência contrafactual. Por exemplo, a de que as causas devem ser "desviantes" ou anormais: o comportamento do jardineiro era desviante (ele supostamente deveria regar minhas flores, ou ao menos normalmente o faz), mas o da rainha era inteiramente normal (ela nunca rega minhas flores, nem alguém esperaria que ela o fizesse). Ou podemos acomodar o julgamento recorrendo a fatores pragmáticos, como a relevância: as omissões do jardineiro e da rainha são, igualmente, causas de *e*, mas na maioria dos contextos de conversação, a omissão da rainha simplesmente não é relevante. Por exemplo, se estou interessado em descobrir quem é o culpado pela morte das flores, a negligência da rainha é simplesmente irrelevante para minha investigação, uma vez que é óbvio que ela não tem culpa.

O próprio Lewis adota essa última abordagem invariantista (Lewis, 2004b). As causas de qualquer evento são muitas e variadas e, em qualquer contexto explicativo, a maioria delas deixa de se salientar; é por isso que nos parece errado mencioná-las (Lewis, 1986d). A abordagem de Lewis explica (alguns) dos fenômenos que motivam o contextualismo, apelando para uma difundida história griceana sobre implicatura conversacional. Uma vantagem do invariantismo é que ele nos permite conceber a causa como uma relação totalmente objetiva, independente da mente — ou, como Menzies (2009: 342) coloca, uma relação "natural".

Não está claro, entretanto, que todos os casos alegados de dependência contextual possam ser tratados dessa maneira. O roubo do bolo de coco da loja por Suzy causou sua doença subsequente: depois de adquirir o bolo, ela o comeu — mas (como logo descobriu) é alérgica ao coco. No entanto, o roubo causou sua doença? Ao que parece, depende daquilo que contrastarmos com o roubo do bolo de coco. Se ela tivesse deixado a loja de mãos vazias — ou roubado um bolo de chocolate — ela não teria ficado doente. Mas, se ela tivesse pagado pelo bolo, em vez de roubá-lo, ainda teria ficado doente. Podemos marcar o contraste pretendido,

por exemplo, pela ênfase: o roubo de Suzy do *bolo de coco* lhe causou a doença, mas seu *roubo* do bolo de coco não.

A teoria original dos eventos de Lewis (1986b) foi feita sob medida para lidar com esses casos — ou ao menos é o que parece. Segundo essa teoria, um evento é um conjunto de regiões espaço-temporais de mundos. Podemos distinguir entre, por exemplo, o evento que é essencialmente o roubo de um bolo de Suzy (e_1) e o evento que é essencialmente ela adquirir (de um jeito ou de outro) um bolo de coco (e_2): os dois eventos consistem em dois diferentes (mas sobrepostos) conjuntos de regiões espaço-temporais de mundos que compartilham seus membros do mundo-atual, ou seja, o que atualmente aconteceu na confeitaria. E assim — pelo menos em face disso — podemos dizer que e_2 foi a causa de sua doença, mas e_1 não foi (uma vez que se não tivesse roubado um bolo, ela teria comprado o bolo de coco).

Não está claro, no entanto, que o apelo às características essenciais dos eventos lide com o problema de modo bem-sucedido. Afinal, e se, se Suzy não tivesse roubado um bolo, o bolo que ela tivesse comprado fosse um bolo de chocolate, não o bolo de coco que ela realmente roubou? (ela de fato queria um bolo, mas não tinha dinheiro suficiente para o bolo de coco). Em qualquer caso, a visão oficial do próprio Lewis é que ao imaginarmos uma suposta causa *c* ausente, nós "imaginamos que *c* é completamente extirpada e limpa da história, sem deixar nenhum fragmento ou aproximação de si mesma" (2004a: 90). Portanto, parece que não conseguimos recuperar a verdade da afirmação de que o *roubo* do bolo de Suzy não foi uma causa de sua doença subsequente. Além disso, a teoria da causação de 2000 de Lewis, no que diz respeito à influência, abandona a distinção entre as essências dos eventos para os quais a resposta acima apelou: temos várias alterações do roubo do bolo de coco (*c*) — incluindo a compra de um bolo de coco e o roubo de um bolo de chocolate, por exemplo — alguns dos quais resultariam em uma alteração do efeito *e* (doença de Suzy) e alguns dos quais não resultariam. O grau de influência de *c* em *e* pode ser ou não suficiente para fazer com que *c* fosse uma causa de *e*; de qualquer forma, "o roubo do bolo de coco por Suzy foi uma causa da sua doença" resulta verdadeiro ou falso independentemente do contexto, o que — de acordo com a ou o contextualista — é o resultado errado (a ou o invariantista, no entanto, pode insistir que não há nenhum problema real aqui. "Porque ela roubou um bolo de coco" seria uma resposta inadequada à pergunta "por que Suzy está doente?", se o pedido vier do médico, que não está interessado em como ela adquiriu o bolo; mas seria uma resposta apropriada no contexto de uma discussão sobre, digamos, Suzy recebendo punição por seu hábito de furtar lojas).

Cei Maslen (2004), Jonathan Schaffer (2005) e Robert Northcott (2008) defendem explicações de causação por "contrastividade". Schaffer concebe a causação como uma relação quaternária — *c,* em vez de *c**, causou *e,* em vez de *e** — e afirma que o contexto (ou outros artifícios, como a ênfase em uma palavra específica) geralmente corrige o valor dos contrastes implícitos (*c** e *e**) em nossa conversa comum, que concebe a relação causal como binária, desempenhando assim um papel na verdade ou falsidade de nossas afirmações causais (binárias). Note que o contrastivismo sobre causação é uma posição distinta da visão de que explicações são (sempre ou algumas vezes) contrastantes (ver, por exemplo, Lewis, 1986d, §VI; Lipton, 1991; Hitchcock, 1999). Em uma visão contrastivista da explicação, as explicações (sempre ou às vezes) assumem a forma "Por que *P* em vez de *Q*?", onde o contraste (*Q*) pode ser explicitamente declarado ou inferido pelo contexto em que a pergunta "Por que *P*?" é feita. Tal visão é totalmente compatível com uma visão invariantista de causação, já que o papel do contraste pode ser apenas selecionar qual das causas de *P* é citada adequadamente ao responder a questão. Observe também que o contrastivismo sobre a explicação não parece resolver o (alegado) problema de modo simples. No caso do roubo do bolo por Suzy, é o contraste do lado das causas (e, portanto, o *explananda*) em que isso está em questão, e não o contraste do lado do efeito (*explanandum*); não está claro como podemos variar o contraste contextualmente relevante para "Suzy adoeceu" de tal modo que diferentes contrastes produzam diferentes vereditos sobre se "Suzy roubou o bolo de coco" é um *explanans* apropriado.

Embora a descrição contrastivista da causação seja geralmente considerada uma versão do contextualismo causal, o contrastivismo é, no entanto, invariantista, no sentido de que existe um fato independente do contexto da questão sobre a qual as relações causais quaternárias se sustentam: de fato, é verdade que Suzy ficou doente, e não permaneceu sadia, por ter roubado o bolo em vez de ter saído de mãos vazias; e falso que ela ficou doente, e não permaneceu sadia, por ter roubado o bolo, em vez de tê-lo comprado. A explanação contrastivista pode, portanto, ser vista como uma espécie de posição intermediária entre invariantismo e contextualismo (veja Steglich-Petersen, 2012; e Montminy & Russo, 2016, para discussões críticas de contextualismo/contrastivismo).

O debate contextualista/invariantista surge de uma forma ligeiramente diferente em grande parte do debate sobre a modelagem de equações estruturais (ver §5.5 abaixo), para a qual nos voltaremos agora.

5. A Modelagem de Equações Estruturais

Várias filósofas contemporâneas e vários filósofos contemporâneos exploraram uma abordagem contrafactual alternativa para a causação que emprega a modelagem de equações estruturais (os primeiros expoentes incluem Hitchcock 2001, 2007; Woodward, 2003; Woodward e Hitchcock, 2003). Essa modelagem, que tem sido usada na área das ciências sociais e biomédicas desde as décadas de 1930 e 1940, recebeu sua formulação mais avançada no livro histórico de 2000 de Judea Pearl. Hitchcock e Woodward reconhecem sua dívida com o trabalho de Pearl e com o trabalho relacionado a redes bayesianas causais de Peter Spirtes, Clark Glymour e Richard Scheines (2001). Contudo, enquanto Pearl e Spirtes, Glymour e Scheines se concentram em questões relacionadas à descoberta e inferência causais, Woodward e Hitchcock se concentram em questões do significado de afirmações causais. Por esse motivo, suas formulações da modelagem de equações estruturais são mais adequadas aos propósitos desta discussão. A exposição desta seção segue em grande parte a de Hitchcock (2001).

5.1. MEE: Introdução Básica

A modelagem de equações estruturais descreve a estrutura causal de um sistema em termos de um modelo causal do sistema, o qual é identificado como um par ordenado <*V*, *E*>, em que *V* é um conjunto de variáveis e *E* um conjunto de equações estruturais que estabelece relações determinísticas entre as variáveis (vamos limitar nossa atenção a sistemas determinísticos aqui; veja §5.4 para uma breve discussão sobre causação indeterminada). As variáveis em *V* descrevem os diferentes estados possíveis do sistema em questão. Embora elas possam assumir qualquer quantidade de valores, nos simples exemplos a serem considerados aqui, as variáveis são binárias e assumem o valor 1, se algum evento ocorre, e 0 se o evento não ocorrer. Por exemplo, formulemos um modelo causal para descrever o sistema anteriormente exposto acerca da preempção tardia, relacionada ao arremesso de pedras de Billy e Suzy. Podemos descrever o sistema usando o seguinte conjunto de variáveis:

- *BT* = 1 se Billy joga uma pedra, e 0 caso contrário;

- *ST* = 1 se Suzy joga uma pedra, e 0 caso contrário;
- *BH* = 1 se a pedra de Billy atingir a garrafa, e 0 caso contrário;
- *SH* = 1 se a pedra de Suzy atingir a garrafa, e 0 caso contrário;
- *BS* = 1 se a garrafa se quebrar, e 0 caso contrário.

Aqui as variáveis são binárias. Mas um modelo diferente pode usar variáveis multivaloradas para representar as diferentes maneiras pelas quais Billy e Suzy jogaram suas pedras, se suas pedras atingiram a garrafa ou se a garrafa se quebrou.

As equações estruturais em um modelo especificam quais variáveis devem ser mantidas fixas em seus valores atuais, e como os valores de outras variáveis dependem uns dos outros. Existe uma equação estrutural para cada variável. A forma tomada por uma equação estrutural para uma variável depende de qual tipo de variável ela é. A equação estrutural para uma *variável exógena* (cujos valores são determinados por fatores externos ao modelo) leva a forma de $Z = z$, que simplesmente indica o valor atual da variável. A equação estrutural para uma *variável endógena* (cujos valores são determinados por fatores internos ao modelo) indica como o valor da variável é determinado pelos valores das outras variáveis. Essa assume a forma:

$$Y = f (X_1, \ldots, X_n)$$

O que essa equação estrutural significa? De fato, existem interpretações concorrentes. Pearl (2000) considera as equações estruturais como conceitos primitivos de sua modelagem, descrevendo-as como representando “os mecanismos básicos” do sistema sob investigação. No entanto, para fins de exposição, é mais conveniente seguir a interpretação de Woodward (2003) e Hitchcock (2001), que pensam nas equações estruturais como expressando certas contrafactuais básicas da seguinte forma:

Se fosse o caso que $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, então seria o caso que $Y = f (x_1, \ldots, x_n)$

Como sugere essa forma de contrafactual, as equações estruturais devem ser lidas da direita para a esquerda: o antecedente das contrafactuais afirma valores possíveis das variáveis X_1 a X_n e o consequente afirma o valor correspondente da

variável endógena *Y*. Existe uma contrafactual desse tipo para cada combinação de valores possíveis das variáveis X_1 a X_n. É importante notar que uma equação estrutural desse tipo não é, portanto, estritamente falando, uma identidade, pois há uma assimetria da direita para a esquerda embutida. Essa assimetria corresponde à assimetria das contrafactuais irretroativas. Por exemplo, supondo que a situação real seja a que nem Suzy nem Billy joga uma pedra a fim de não quebrar a garrafa. A contrafactual irretroativa "Se Suzy ou Billy tivessem jogado uma pedra, a garrafa teria se quebrado" é verdadeira. Mas a contrafactual "Se a garrafa tivesse quebrado, Suzy ou Billy teriam jogado uma pedra" é falsa.

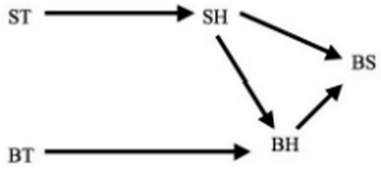
Para ilustrar, considere o conjunto de equações estruturais que podem ser usadas para modelar o exemplo de preempção tardia de Billy e Suzy. Dado o conjunto de variáveis *V* listado acima, os membros do conjunto das equações estruturantes *E* podem ser colocados da seguinte forma:

- *ST = 1;*
- *BT = 1;*
- *SH = ST;*
- *BH = BT & ~ SH;*
- *BS = SH v BH.*

Nessas equações, símbolos lógicos são usados para representar funções matemáticas em variáveis binárias: $\sim X = 1 - X$; $X \vee Y = max\{X, Y\}$; $X \& Y = min\{X, Y\}$. As duas primeiras equações simplesmente indicam os valores reais das variáveis exógenas *ST* e *BT*. A terceira equação codifica duas contrafactuais, uma para cada valor possível de *ST*. Afirma-se que, se Suzy tivesse jogado uma pedra (o que de fato ela fez), ela teria atingido a garrafa; e se ela não tivesse jogado uma pedra, não teria atingido a garrafa. A quarta equação codifica quatro contrafactuais, uma para cada combinação possível de valores para *BT* e ~ *SH*. Afirma-se que, se Billy tivesse jogado uma pedra e a pedra de Suzy não tivesse atingido a garrafa, a pedra de Billy teria atingido a garrafa, mas não o teria se uma ou mais dessas condições não tivessem sido atendidas. A quinta equação também codifica quatro contrafactuais, uma para cada combinação possível de valores para *SH* e *BH*. Afirma-se que, se a pedra de Suzy ou a pedra de Billy (ou possivelmente ambas) tivesse atingido a garrafa, ela teria se quebrado; mas se nenhuma pedra tivesse atingido a garrafa, esta não se quebraria.

As equações estruturais acima podem ser representadas em termos de um *gráfico direcionado*. As variáveis no conjunto *V* são representadas como nós no gráfico. Uma seta direcionada de um nó *X* para outro *Y* representa o fato de que a variável *X* aparece no lado direito da equação estrutural para *Y*. Nesse caso, diz-se que *X* é antecedente de *Y*. Variáveis exógenas são representadas por nós quando não possuem setas direcionadas a elas. Um *caminho direcionado* de *X* a *Y* em um gráfico é uma sequência de setas que conecta *X* a *Y*. O gráfico direcionado do modelo do exemplo descrito acima sobre o exemplo de Billy e Suzy está representado na Figura 1 abaixo:

Figura 1.



As setas nessa figura nos dizem que o despedaçar da garrafa é uma função das pedras de Suzy e Billy atingindo a garrafa; que a pedra de Billy atingindo a garrafa é uma função de Billy atirando uma pedra e a pedra de Suzy atingindo a garrafa; e que a pedra de Suzy atingindo a garrafa é uma função dela atirando a pedra. É importante observar que os nós no gráfico representam variáveis e não — como no caso de "diagramas de neurônios", encontrados em Lewis (1986c) — valores de variáveis. Observa-se também que, embora as setas nos digam que existem relações de dependência contrafactual entre os valores das variáveis — todos os valores possíveis, não apenas os atuais —, não nos dizem quais são essas relações de dependência; para isso, é preciso olhar para as equações estruturais. Por exemplo, o gráfico direcionado nos diz apenas que o valor de *BH* depende contrafactualmente, de uma forma ou de outra, dos valores de *BT* e *SH* (isso não nos diz, por exemplo, que se Billy tivesse atirado a pedra e a pedra de Suzy não tivesse atingido a garrafa, a pedra de Billy a teria atingido).

Como vimos, as equações estruturais codificam diretamente algumas contrafactuais. No entanto, algumas contrafactuais que não são diretamente codificadas podem ser derivadas delas. Considere, por exemplo, a contrafactual "Se a pedra de Suzy não tivesse atingido a garrafa, ela ainda assim teria se quebrado".

De fato, a pedra de Suzy atingiu a garrafa. Mas podemos determinar o que teria acontecido caso não tivesse atingido, substituindo a equação estrutural da variável endógena *SH* pela equação *SH* = 0, mantendo todas as outras equações inalteradas. Portanto, em vez de ter seu valor determinado de modo usual através da variável *ST*, o valor de *SH* é definido "milagrosamente". Pearl descreve isso como uma "intervenção cirúrgica" que altera o valor da variável. Em termos de representação gráfica, isso significa apagar a seta da variável *ST* para a variável *SH* e tratar a *SH* como se fosse uma variável exógena. Após essa operação, o valor da variável *BS* pode ser calculado e apresentado como igual a 1: dado que Billy havia jogado sua pedra, sua pedra teria atingido a garrafa e a quebrado. Portanto, essa contrafactual em particular é verdadeira. Esse procedimento para avaliar contrafactuais tem afinidades diretas com a interpretação de contrafactuais irretroativas de Lewis: a intervenção cirúrgica que define a variável *SH* em seu valor hipotético, mas mantém todas as outras equações inalteradas, é semelhante em seus efeitos ao pequeno milagre de Lewis, que realiza o antecedente da contrafactual, mas preserva o passado.

Em geral, para avaliar uma contrafactual, digamos "Se X_1, ..., X_n forem o caso, então ...", substitui-se a equação original para cada variável X_i por uma nova equação estipulando seu valor hipotético, ao mesmo tempo em que se mantém as outras equações inalteradas; então se calcula os valores para as variáveis restantes, de modo a verificar se elas tornam o consequente verdadeiro. Essa técnica de substituir uma equação por um valor hipotético definido por uma "intervenção cirúrgica" permite capturar a noção de dependência contrafactual entre variáveis:

> (8) Uma variável *Y depende contrafactualmente* de uma variável *X* em um modelo se, e somente se, for realmente o caso que $X = x$ e $Y = y$ e existirem valores $x' \neq x$ e $y' \neq y$ de modo que, substituindo-se a equação por *X* com $X = x'$, produz-se $Y = y'$.

Obviamente, até agora temos apenas algo que estamos chamando de "modelo causal", <*V*, *E*>; não nos foi dito nada sobre como extrair informações causais disso. Como já deve estar claro, a receita básica será, aproximadamente, a seguinte: a verdade de "*c* causa *e*" (ou "*c* é uma causa real de *e*"), onde *c* e *e* são eventos-espécime particulares, será um caso de relação contrafactual, conforme codificada pelo modelo, entre duas variáveis *X* e *Y*, quando a ocorrência de *c* é representada por uma equação estrutural da forma $X = x_1$ e a ocorrência de *e* é representada por uma equação estrutural da forma $Y = y_1$. No entanto, podemos

ver imediatamente que não podemos identificar a causação diretamente através da dependência contrafactual, conforme definido em (8) acima. Teríamos a verdade de "O arremesso de Suzy fez sua pedra acertar a garrafa" (*ST* = 1 e *SH* = 1 e, como *SH* = *ST* é um membro de *E*, sabemos que se substituirmos *ST* = 1 por *ST* = 0, obteremos *SH* = 0). Mas não teríamos, por exemplo, a verdade de "O arremesso de Suzy fez com que a garrafa se quebrasse", pois se substituirmos *ST* = 1 por *ST* = 0 e trabalharmos com as equações, ainda terminaremos com *BS* = 1.

Como, então, podemos definir "causação efetiva"[11] usando a estrutura de equações estruturais? Para isso, consideraremos como a MEE lida com os casos de preempção tardia, como o caso de Suzy e Billy. Halpern e Pearl (2001, 2005), Hitchcock (2001) e Woodward (2003) fornecem aproximadamente o mesmo tratamento para a preempção tardia. A maneira encontrada por eles é o emprego de certo procedimento para testar a existência de uma relação causal. O procedimento é procurar um processo intrínseco que conecte supostamente a causa e o efeito; suprimir outras influências não intrínsecas "congelando" o seu entorno como ele realmente é; e depois submeter a suposta causa a um teste contrafactual. Portanto, por exemplo, para testar se o arremesso da pedra por Suzy causou a quebra da garrafa, deveríamos examinar a execução do processo de *ST* através de *SH* para *BS*; manter fixa em seu valor atual (ou seja, 0) a variável *BH*, que é extrínseca a esse processo; e, então, modificar a variável *ST* para ver se ela altera o valor de *BS*. Os últimos passos envolvem avaliar a contrafactual "Se Suzy não tivesse jogado uma pedra e a pedra de Billy não tivesse atingido a garrafa, a garrafa não teria se quebrado". É fácil ver que essa contrafactual é verdadeira. Por outro lado, quando realizamos um procedimento semelhante para testar se Billy, ao ter jogado uma pedra, fez com que a garrafa se quebrasse, somos obrigados a considerar a contrafactual "Se Billy não tivesse jogado sua pedra e a pedra de Suzy tivesse atingido a garrafa, a garrafa não teria se quebrado". Essa contrafactual é falsa. É a diferença nos valores de verdade dessas duas contrafactuais que explica o fato de que foi o lançamento de pedra de Suzy, e não o de Billy, que causou a quebra da garrafa (uma teoria semelhante é desenvolvida em Yablo (2002, 2004), embora não no modelo de equações estruturais).

---

[11] N.T.: Traduzimos o termo "*actual causation*" como "causação efetiva", mas note que a noção de *efetividade* aqui empregada não deve ser confundida com uma noção temporal.

Hitchcock (2001) apresenta uma organização útil desse raciocínio. Ele define uma *rota* entre duas variáveis *X* e *Z* no conjunto *V* como sendo uma sequência ordenada de variáveis < *X*, Y_1, …, Y_n, *Z* > de modo que cada variável na sequência esteja em *V* e seja um genitor de suas variáveis sucessoras na sequência. Uma variável *Y* (distinta de *X* e *Z*) é *intermediária* entre *X* e *Z* se, e somente se, pertencer a alguma rota entre *X* e *Z*. Em seguida, ele introduz o novo conceito de uma rota causal ativa:

> (9) A rota < *X*, Y_1, …, Y_n, *Z* > é *ativa* no modelo causal < *V*, *E* > se, e somente se, *Z* depende contrafactualmente de *X* no novo sistema de equações *E'* construído a partir de *E* como a seguir: para todo *Y* em *V* , se *Y* for intermediário entre *X* e *Z*, mas não pertencer à rota <*X*, Y_1, …, Y_n, *Z*>, então substitua a equação de *Y* por uma nova equação que defina *Y* igual ao seu valor atual em *E* (se não houver variáveis intermediárias que não pertençam a essa rota, então *E'* será apenas *E*). (Hitchcock, 2001: 286).

Essa definição generaliza a ideia informal esboçada no exemplo de Suzy e Billy. < *ST*, *SH*, *BS* > é uma rota causal ativa, porque quando mantemos *BH* fixa em seu valor atual (a pedra de Billy não acerta a garrafa), *BS* depende contrafactualmente de *ST*. Por outro lado, a rota < *BT*, *BH*, *BS* > não é ativa, porque quando mantemos *SH* fixa em seu valor atual (a pedra de Suzy bate na garrafa), *BS* não depende contrafactualmente de *BT*.

Em termos da noção de uma rota causal ativa, Hitchcock define a causação efetiva (ou causação-espécime) nos seguintes termos:

> (10) Se *c* e *e* são eventos atuais distintos, e *X* e *Z* são variáveis binárias cujos valores representam a ocorrência e a não ocorrência desses eventos, então *c* é uma causa de *e* se, e somente se, houver uma rota causal ativa de *X* a *Z* em um modelo causal apropriado < *V*, *E* >.

(Voltaremos à noção de um "modelo causal apropriado" em §5.3 abaixo).

Do modo como foi exposto, (10) não lida com casos de sobredeterminação simétrica — como quando Suzy e Billy atiram suas pedras de forma independente, cada lançamento sendo suficiente para a garrafa se quebrar e ambas as pedras batendo em seguida na garrafa — de modo que nenhum lançamento é preemptivo em relação ao outro, já que nenhum dos lançamentos está em uma rota ativa,

conforme definido em (9). Para lidar com esses casos, Hitchcock enfraquece (10) substituindo a "rota ativa" em (10) pela noção de uma *rota fracamente ativa* (Hitchcock, 2001: 290). A ideia essencial aqui é a de que existe uma rota fracamente ativa entre *X* e *Z*, apenas quando *Z* depende contrafactualmente de *X* sob o congelamento de alguns valores possíveis, não necessariamente atuais, das variáveis que não estão na rota de *X* para *Z*. Intuitivamente, para recuperar a dependência contrafactual entre o arremesso de Suzy e a quebra, mantemos *BT* = 0 fixo: se Suzy não tivesse arremessado, *no modelo em que Billy não arremessa,* a garrafa não teria se quebrado. Da mesma maneira ocorre com o arremesso de Billy.

A estratégia básica implantada aqui, para lidar tanto com a preempção e a sobredeterminação simétrica, repousa em uma semelhança óbvia com a solução de quase-dependência de Lewis para o problema da preempção tardia. Lewis recorre à quase-dependência porque a quebra da garrafa (*e*) não depende contrafactualmente do arremesso de Suzy (*c*), graças ao que teria acontecido se ela não tivesse arremessado (nomeadamente*,* a pedra de Billy que teria quebrado a garrafa); *e* quase-depende de *c*, no entanto, devido ao fato de que em um mundo possível com as mesmas leis, no qual o caráter intrínseco do processo de *c* para *e* seja o mesmo, mas em que Billy não atira, há a dependência contrafactual requerida. Ao "congelarmos" variáveis que não são intrínsecas ao processo *c*-*e* em seus valores atuais (em casos de preempção tardia) — por exemplo, congelarmos *BH* em 0 — temos aproximadamente o mesmo artifício. A principal diferença é que a solução de Lewis envolve apelar para a verdade de uma contrafactual perfeitamente comum ("Se Suzy não atirou, ...") em um mundo possível onde alguns eventos atuais (por exemplo, a pedra de Billy atingir) não ocorrem, enquanto que a solução de equações estruturais envolve apelar para a verdade de uma contrafactual com um tipo especial de antecedente ("Caso Suzy não tivesse atirado e a pedra de Billy também não atingisse, …"). Hitchcock chama essas contrafactuais de "explicitamente imprevisíveis"[12] (ENF). (Da mesma forma para a sobredeterminação simétrica, em que "congelamos" *BT* em 0 — desta vez um valor não atual — para recuperar a dependência contrafactual entre o arremesso de Suzy e a quebra).

---

[12] N.T.: Traduzimos a expressão "*explicitly nonforetracking*" como "explicitamente imprevisíveis", dada a carência de um termo em português mais preciso para a tradução de "*nonforetracking*".

5.2. MEE e Contrafactuais

Aquelas e aqueles que adotaram a abordagem da Modelagem de Equações Estruturais (MEE) para fornecer uma análise da causação "efetiva" — isto é, a relação causal entre eventos atuais e particulares — tiveram muito pouco a dizer sobre a semântica das contrafactuais que sustentam o MEE. Algumas autoras e alguns autores (*e.g.*, Hitchcock, 2001) explicitamente — ou implicitamente — assumem uma abordagem amplamente lewisiana das contrafactuais, de modo que as equações estruturais são representações de relações de fatos sobre a dependência contrafactual — como descrito acima — cujas condições de verdade são extensamente lewisianas.

Por outro lado, pode-se ir à outra direção, usando a abordagem MEE para fornecer as condições de verdade para as contrafactuais, tratando as equações estruturais (como *SH* = *ST*) como representações de relações de dependência *causais*, que por sua vez fornecem tais condições de verdade (Galles & Pearl 1998, Woodward e Hitchcock 2003, Schulz 2011, Briggs 2012). De maneira semelhante, pode-se evitar uma concepção lewisiana, baseada em uma intervenção de milagres, e definir intervenções em termos explicitamente causais (ver, por exemplo, Woodward e Hitchcock 2003: 12–13).

A escolha entre essas duas maneiras diferentes de proceder se conecta ao debate mais amplo sobre se a causa deve ser analisada em termos de contrafactuais ou vice-versa. Lewis, é claro, adota a primeira abordagem. A escolha de fazê-lo — pelo menos para ele — se dá por se encaixar dentro de uma agenda amplamente humiana: uma vez que a causa é uma noção modal, ela ameaça a tese da sobreveniência humiana (Lewis, 1986a, ix), a menos que, de alguma forma, ela pudesse ser descrita em termos de relações de semelhança entre mundos, onde essas relações de semelhança não recorram, por sua vez, às características causais (ou outras que violam a sobreveniência humiana) dos mundos. A análise de contrafactuais de Lewis, juntamente com a sua análise de leis, têm esse papel. Em contrapartida, outras autoras e outros autores argumentaram que esse artifício simplesmente não pode ser feito dessa maneira: não podemos analisar contrafactuais sem apelar à causação (Edgington, 2011).

Há questões metafísicas profundas em jogo, portanto: pode-se ver a abordagem MEE como oferecendo uma variante mais sofisticada da abordagem de Lewis, a qual compartilha as aspirações reducionistas desta última. Ou se pode — especialmente se estivermos céticos sobre as perspectivas dessas aspirações reducionistas — adotar a abordagem MEE com espírito antirreducionista, vendo-a

não como uma maneira de definir causação em termos não-causais, mas, em vez disso, como uma maneira de extrair informações causais úteis e sofisticadas a partir de um modelo inerentemente causal de uma dada situação complexa.

5.3. Modelos e Realidade

É uma característica geral da abordagem MEE que o modelo não precisa incluir como variáveis todos os fatores relevantes para o efeito em consideração (e, de fato, nenhum modelo inclui — pois existem fatores demais). No modelo de Billy/Suzy acima, por exemplo, não há variáveis que descrevem os estados atuais e possíveis correspondentes aos intermediários causais entre os arremessos de Billy ou Suzy (ou da ausência de seus arremessos) e suas respectivas pedras atingindo (ou não) a garrafa. Então, o que determina quais variáveis devem ou não ser incluídas no modelo, a fim de descobrirmos as relações causais entre as variáveis nas quais estamos interessadas e interessados?

É importante enfatizar que não existe um modelo exclusivamente correto para ser adotado em uma determinada situação. Um modelo que, por exemplo, interpolasse um grande número de intermediários entre o arremesso de Suzy e o choque de sua pedra na garrafa, revelaria mais sobre a estrutura causal da situação atual e de várias alternativas contrafactuais diferentes. Mas isso não o torna o modelo "certo" para considerar o estado causal dos respectivos lançamentos, de Billy e Suzy, em relação ao despedaçar da garrafa. Tal modelo produziria o mesmo resultado que o mais simples descrito anteriormente e, portanto, as variáveis adicionais seriam simplesmente uma complicação desnecessária. Por outro lado, existem limites para o que podemos deixar de fora. Por exemplo, um modelo causal que inclui apenas *ST* e *BS* como variáveis não produziria o resultado de que o lance de Suzy causou a quebra da garrafa, pois essa contrafactual não é verdadeira nesse modelo (para que isso se torne realidade, precisamos incluir *BH* e mantê-lo fixo em seu valor atual, *BH* = 0).

Então, quais são as restrições nos modelos causais, de modo que eles representem com precisão os fatos causais nos quais estamos interessadas e interessados (Halpern e Hitchcock, 2010: §§4-5)? Várias autoras e vários autores propuseram restrições que nos dizem o que contam como (para usar o termo de Hitchcock) modelos “aptos”, muitas das quais são análogas às restrições de Lewis e pelas mesmas razões, nomeadamente, para garantir que não haja dependência

contrafactual espúria. Assim, Hitchcock (2001: 287) propõe que os valores das variáveis não deveriam representar eventos que tenham relações lógicas ou metafísicas entre si, e Blanchard e Schaffer (2017: 182) propõem que os valores atribuídos deveriam representar caracterizações intrínsecas. Hitchcock (2001: 287) também propõe que as variáveis não deveriam receber valores "que não se deseja levar a sério" (mais sobre isso a seguir). Halpern e Hitchcock (2010) adicionam uma restrição de “estabilidade”: o acréscimo de variáveis adicionais não deve anular os veredictos causais (essa restrição aborda o problema do “modelo” descrito acima que inclui apenas *ST* e *BS*; esse modelo fornece um veredicto, a saber, que o arremesso de Suzy não quebra a garrafa, o que é revertido pela adição de variáveis adicionais). E Hitchcock (2007: 503) propõe a restrição de que o modelo “inclua variáveis suficientes para capturar a estrutura essencial da situação que está sendo modelada” (entretanto, se tivéssemos aspirações reducionistas, essa restrição pareceria tornar a análise da causação viciosamente circular, já que a "estrutura essencial" da situação é presumivelmente sua estrutura *causal* essencial — exatamente o que um modelo causal deve fornecer).

Precisamente, quais deveriam ser as restrições nos modelos “aptos” ou “apropriados” é uma questão de debate filosófico em curso (Blanchard e Schaffer, 2017: §1.3). O foco aqui está nas restrições que garantem que um modelo não produz resultados espúrios (*e.g.*, que o arremesso de Suzy não faz com que a garrafa se quebre). No entanto, o MEE também é usado como uma ferramenta prática na investigação científica e isso coloca em jogo questões normativas adicionais sobre a escolha de variáveis e sua faixa de valores permitidos. No contexto de se atribuir culpa pela garrafa quebrada, por exemplo, não seria relevante incluir a resistência do vidro da garrafa como variável; por contraste, um dono de uma loja local — cansado do vandalismo de Suzy e Billy (eles rotineiramente quebram a vitrine de sua loja e as garrafas que encontram) — pode estar muito interessado na questão de que a resistência do vidro seja suficiente para suportar o arremesso de pedras (veja, por exemplo, Woodward (2016), Hitchcock (2017) e — para um exemplo prático no contexto de inferência causal no aprendizado de máquinas — Chalupka, Eberhardt & Perona (2017)).

5.4. MEE e Causação Incerta

Como vimos em §1.4 anteriormente, Lewis revisou sua explanação de causação de 1973 para levar em conta a causação incerta. Qualquer explicação da causação que se baseia na ideia de que as causas aumentam as chances de seus efeitos encontra dois problemas principais: o aumento da chance não é necessário, nem mesmo suficiente para a causa (Caso 1: O médico reduz a chance muito alta de um ataque cardíaco do paciente através de uma cirurgia drástica. Infelizmente, a cirurgia em si faz com que o paciente tenha um ataque cardíaco. A cirurgia diminui a chance, mas causa o ataque cardíaco. Caso 2: Billy e Suzy estão arremessando pedras na garrafa novamente. Cada um dos seus arremessos aumenta a chance da garrafa se quebrar, mas o arremesso de Suzy antecipa o de Billy. O arremesso de Billy aumenta a chance, mas não causa o despedaçar da garrafa).

O primeiro tipo de caso pode (talvez) ser tratado pela explanação modificada de Lewis de 1973, ao encontrar algum evento intermediário *d*, tal que a cirurgia aumenta a chance de *d* e *d*, por sua vez, aumenta a chance de ataque cardíaco. Mas não pode lidar com o segundo tipo de caso: o arremesso de Billy atende à condição suficiente de Lewis para a causação incerta e, portanto, a explanação modificada de 1973 o considera erroneamente como uma causa (Menzies, 1996). Esse é um problema que Lewis viu por conta própria e nunca solucionou: sua explicação posterior de causação por "influência", de 2000, assume o determinismo (Lewis, 2000: n.1; 2004a: 79-80) e, portanto, ignora o problema. Exemplos de ambos os tipos têm sido objeto de extensa discussão no contexto de teorias contrafactuais e probabilísticas da causação (para discussões sobre a melhor forma de lidar com os exemplos, dentro de teorias que não assumem determinismo, consulte Barker, 2004; Beebee, 2004a; Dowe, 2000, 2004; Hitchcock, 2004; Kvart, 2004; Noordhof, 1999, 2004; Ramachandran, 1997, 2004).

Do mesmo modo, as explicações do MEE majoritariamente assumem o determinismo: como na descrição original de Lewis de 1973, o alicerce básico de tais descrições é a dependência contrafactual não-incerta. No entanto, houve tentativas recentes de estender as análises baseadas em MEE para abranger a causação incerta por Fenton-Glynn (Glynn, 2011; Fenton-Glynn, 2017; ver também o item §4.4 em *Causação Probabilística*[13]).

---

[13] Em https://plato.stanford.edu/archives/spr2021/entries/causation-probabilistic/

5.5. Padrões e Desvios

Em §4, vimos dois exemplos dos tipos de casos que motivaram alguns autores a endossar o contextualismo. Os exemplos apresentam diferentes características. No caso do roubo do bolo de coco, a ideia é que, em diferentes contextos de enunciado, a mesma alegação causal — "o roubo do bolo de coco da loja por Suzy causou sua doença subsequente" — pode variar em valor de verdade, dependendo se o contexto determina que o que está em questão é a criminalidade de Suzy (roubo, em vez de compra), ou qual item ela roubou (bolo de coco em vez de bolo de chocolate). O caso do jardineiro e da rainha, por outro lado, é o caso em que a (alegada) causa (omissão do jardineiro) e a não-causa (da rainha) mantêm a mesma relação contrafactual com o efeito e, ainda assim, os julgamentos diferem com relação à sua condição causal.

O foco principal do debate dentro da abordagem MEE tem sido o segundo tipo de caso — casos nos quais Menzies chama de "contrafactuais isomórficos" (Menzies, 2017), onde dois cenários diferentes têm modelos causais isomórficos e, ainda assim, nossos julgamentos sobre o que causa o que diferem em ambos os cenários. Por exemplo, considere um caso de "falsa prevenção" (Hiddleston, 2005): acreditando que o assassino já havia envenenado o café da vítima, o guarda-costas coloca um antídoto nele. No entanto, o assassino mudou de ideia e não envenenou o café. A vítima sobrevive — mas sua sobrevivência não foi, certamente, causada pela ação do guarda-costas, já que não havia ameaça à vida da vítima que o guarda-costas tivesse neutralizado. É possível, no entanto, construir um modelo causal deste caso que seja isomórfico para um caso padrão de sobredeterminação simétrica: Billy e Suzy estão atirando pedras novamente, mas desta vez os dois arremessos atingem a garrafa ao mesmo tempo, de modo que cada um é suficiente para quebrá-la e nenhum deles se antecipa ao outro. Nesse caso (supostamente), identificamos os arremessos de Billy e Suzy como causas da quebra da garrafa (Blanchard e Schaffer, 2017: 185-6).

Uma ampla linha de respostas ao problema dos contrafactuais isomórficos tem sido distinguir entre eventos "padrão" (ou "normais") e "desviantes", e criar essa distinção de modo a extrair as informações causais. Por exemplo, a abordagem de Menzies explora o mecanismo visto anteriormente com relação à solução de Hithcock para o problema de sobredeterminação simétrica, o qual envolve a fixação de

variáveis “fora do caminho” em valores não atuais. A sugestão de Menzies é que fixemos essas variáveis nos seus valores "mais normais", para que, efetivamente, avaliemos as contrafactuais relevantes sob a perspectiva de um mundo onde esses valores normais são atualizados, em vez de fazê-lo sob a perspectiva do mundo atual (Menzies, 2004, 2007, 2009).

Intuitivamente, a ideia básica é que (no caso de sobredeterminação) o valor “mais normal” de cada *BT* e *ST* é 0 (atirar pedras em garrafas não é normal!). Portanto, da perspectiva de um mundo onde *ST* = 0, *BS* depende contrafactualmente de *ST* (e similarmente de *BT*). Assim, os arremessos de Billy e Suzy contam como causas da quebra da garrafa. Envenenar o café de alguém também não é normal. Por isso, mantemos a desistência do assassino em envenenar o café fixa (assim, nesse caso, o mundo “mais normal” é exatamente o mundo atual no que diz respeito ao envenenamento), o que oferece o resultado certo de que a sobrevivência da vítima não depende contrafactualmente da administração do antídoto pelo guarda-costas (veja Hitchcock (2007) para uma solução diferente que se baseia na distinção padrão/desvio).

É difícil ver, no entanto, como poderia haver uma noção unívoca e razoavelmente bem definida de “normalidade” a qual daria veredictos claros sobre quais são os valores “normais” ou “padrão” das variáveis e, portanto, forneceria um conjunto de modelos objetivamente “corretos”, todos com o mesmo veredicto para a mesma situação (Blanchard e Schaffer, 2017: §§2 e 3). Blanchard e Schaffer argumentam que a "relatividade padrão" não resolve alguns dos problemas que deveria solucionar; mais importante, ainda, argumentam que os supostos casos de modelos causais isomórficos não são, em primeiro lugar, casos realmente genuínos: eles surgem porque um (ou ambos) dos modelos relevantes não cumpre os critérios independentemente motivados para a “aptidão” (ver §5.3 acima). Por exemplo, um desses critérios é que as variáveis não recebam valores "que ninguém se dispõe a levar a sério" (Hitchcock, 2001: 287). Mas o caso do jardineiro e da rainha viola esse critério: a possibilidade da rainha regar minhas flores é precisamente aquela que não levamos a sério. Em outros casos, eles argumentam, o isomorfismo resulta da implantação de modelos “empobrecidos”: modelos que não incluem variáveis suficientes para representar adequadamente a “estrutura causal essencial” da situação que está sendo modelada (Blanchard e Schaffer, 2017: §3).

A visão de Blanchard e Schaffer é que, na medida em que nossos julgamentos causais exibem algo como a relatividade-padrão, isso se deve a heurísticas cognitivas que geram vieses em nossos julgamentos. Alternativas para eventos "desviantes" tendem a “saltar” para nós — elas se salientam porque são fáceis de imaginar —

enquanto as alternativas para eventos "padrão" não. A visão de Blanchard e Schaffer pode, portanto, ser vista como uma versão do invariantismo, em que os tipos de casos que supostamente deveriam motivar o contextualismo são acomodados por características de nossas conversas e pensamentos causais, os quais são estranhos ao conceito (não-dependente de norma) de causação.

Uma série de questões permanece nessa área. Uma é se devemos ou não exigir, em primeiro lugar, uma resposta exclusivamente "correta" para a pergunta sobre o que conta como valores "normais" ou "padrão" das variáveis; talvez, se, por exemplo, houvesse dimensões diferentes de "normalidade" (verossimilhança estatística, consonância com normas morais ou legais, etc.), deveríamos adotar a ideia de que dois modelos aptos, de uma mesma situação, poderiam oferecer resultados diferentes e igualmente corretos, dependendo de como os valores normais ou padrão são definidos — o que, por sua vez, dependeria do contexto (*e.g.*, a finalidade para a qual o modelo está sendo implantado).

Questões mais gerais que estão no pano de fundo do debate sobre o "relativismo-padrão" incluem se o objetivo do conceito de causação é mais bem abordado por um conceito de causação "igualitário" ou "invariantista" ou, em vez disso, por um que faz com que o conceito empregue considerações normativas. Hitchcock (2017), por exemplo, argumenta que, como nosso interesse no que causa o que é, em essência, um interesse nos tipos de intervenção que trariam os tipos de resultados que desejamos, deveríamos adotar a última. Uma questão ainda mais geral é se existe apenas um conceito de causação, ao qual todas as explanações disponíveis têm ou deveriam ter como objetivo, ou, em vez disso, vários (Hall 2004, McDonnell, 2018). Talvez ainda, causação seja o que Nancy Cartwright (seguindo Neurath) chama de um conceito de "*Ballung"*[14] : "um conceito com limites ásperos, mutáveis e porosos, um congestionamento de ideias e implicações diferentes que podem, em várias combinações, ser colocados à luz para diferentes fins e em diferentes contextos" (2017: 136).

---

[14] N.T.: Resolvemos manter o termo alemão "*ballung*", cunhado pelo filósofo austríaco Otto Neurath. Uma tradução aproximada, em português, seria "agregado" ou "aglomerado".

Referências

ALBERT, D. Z. **Time and Chance**. Boston: MIT Press, 2000.

BARKER, S. Analysing Chancy Causation without Appeal to Chance-raising. *In:* DOWE, P.; NOORDHOF, P. (ed.). **Cause and Chance:** Causation in an Indeterministic World. London: Routledge, 2004, p. 120-137.

BEEBEE, H. Chance-changing Causal Processes. 2004a. *In:* DOWE, P.; NOORDHOF, P. (ed.). **Cause and Chance:** Causation in an Indeterministic World. London: Routledge, 2004, p. 39-57.

BEEBEE, H. Causing and Nothingness. 2004b. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 291-308.

BEEBEE, H.; HITCHCOCK, C.; MENZIES, P. (ed.). **The Oxford Handbook of Causation**. Oxford: Oxford University Press. 2009.

BEEBEE, H.; HITCHCOCK, C.; PRICE, H. (ed.). **Making a Difference**: Essays on the Philosophy of Causation. Oxford: Oxford University Press, 2017.

BENNETT, J. Event Causation: the Counterfactual Analysis. **Philosophical Perspectives**, v. 1, p. 367-386, 1987.

BLANCHARD, T.; SCHAFFER, J. Cause without Default. *In:* BEEBEE, H.; HITCHCOCK, C.; PRICE, H. (ed.). **Making a Difference**: Essays on the Philosophy of Causation. Oxford: Oxford University Press, 2017, p. 175-214.

BRIGGS, R. Interventionist Counterfactuals. **Philosophical Studies**, v. 160, p. 139-66, 2012.

BUCKLE, S. **Hume's Enlightenment Tract**: The Unity and Purpose of An Enquiry Concerning Human Understanding. Oxford: Oxford University Press, 2004.

CARTWRIGHT, N. Can Structural Equations Explain How Mechanisms Explain? *In:* BEEBEE, H.; HITCHCOCK, C.; PRICE, H. (ed.). **Making a Difference**: Essays on the Philosophy of Causation. Oxford: Oxford University Press, 2017, p. 132-152, 2017.

CHALUPKA, K.; EBERHARDT, F.; PERONA, P. Causal Feature Learning: An Overview. **Behaviormetrica**, v. 44, p. 137-64, 2017.

CHENG, P.; NOVICK, L. Causes versus Enabling Conditions. **Cognition**, v. 40, p. 83-120, 1991.

CLARKE, R.; SHEPHERD, J.; STIGALL, J.; REPKO WALLER, R.; ZARPENTINE, C. Causation, Norms, and Omissions: A Study of Causal Judgements. **Philosophical Psychology**, v. 28, p. 279-293, 2015.

COLLINS, J. Preemptive Preemption. **Journal of Philosophy**, v. 97, p. 223-234, 2000.

COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004.

DOWE, P. **Physical Causation**. Cambridge: Cambridge University Press, 2000.

DOWE, P. Chance-lowering Causes. *In:* DOWE, P.; NOORDHOF, P. (ed.). **Cause and Chance:** Causation in an Indeterministic World. London: Routledge, 2004, p. 28-38.

DOWE, P. Absences, Possible Causation, and the Problem of Non-Locality. **The Monist**, v. 92, p. 23-40, 2009.

DOWE, P.; NOORDHOF, P. (ed.). **Cause and Chance:** Causation in an Indeterministic World. London: Routledge, 2004.

EDGINGTON, D. Causation First: Why Causation is Prior to Counterfactuals. In: HOERL, C.; MCCORMACK, T.; BECK, S. R. (ed.). **Understanding Causation**: Issues in Philosophy and Psychology. Oxford: Oxford University Press, 2011, p. 230-241.

ELGA, A. Statistical Mechanics and the Asymmetry of Counterfactual Dependence. **Philosophy of Science**, v. 68, sup., p. 313-324, 2000.

EHRING, D. Causal Relata. **Synthese**, v. 73, p. 319-328, 1987.

FENTON-GLYNN, L. A Proposed Probabilistic Extension of the Halpern and Pearl Definition of 'Actual Cause'. **The British Journal for the Philosophy of Science**, v. 68, p. 1061-1124, 2017.

FRISCH, M. **Inconsistency, Asymmetry and Non-Locality**: Philosophical Issues in Classical Electrodynamics. New York: Oxford University Press, 2005.

FRISCH, M. Causation, Counterfactuals and Entropy. In: PRICE, H.; CORRY, R. (ed.). **Causation, Physics, and the Constitution of Reality**: Russell's Republic Revisited. Oxford: Oxford University Press, 2007.

GALLES, D.; PEARL, J. An Axiomatic Characterization of Causal Counterfactuals. **Foundations of Science**, v. 3, p. 151-82, 1998.

GLYNN, L. A Probabilistic Analysis of Causation. **The British Journal for the Philosophy of Science**, v. 62, p. 343-392, 2011.

HALL, N. Two Concepts of Causation. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 225-276.

HALPERN, J.; PEARL, J. **Causes and Explanations**: A Structural-model Approach – Part I: Causes. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufman, 2001, p. 194-202.

HALPERN, J.; PEARL, J. Causes and Explanations: A Structural-model Approach – Part I: Causes (expanded version). **British Journal for the Philosophy of Science**, v. 56, p. 843-887, 2005.

HALPERN, J.; HITCHCOCK, C. Actual Causation and the Art of Modeling. *In:* DECHTER, R.; GEFFNER, H.; HALPERN J. (ed.). **Heuristics, Probability, and Causality**: A Tribute to Judea Pearl. London: College Publications, 2010, p. 383-406.

HAUSMAN, D. **Causal Asymmetries**. Cambridge: Cambridge University Press, 1998.

HIDDLESTON, E. Causal Powers. **The British Journal for the Philosophy of Science**, v. 56, p. 27-59, 2005.

HILTON, D.; SLUGOSKI, B. Knowledge-Based Causal Attribution: The Abnormal Conditions Focus Model. **Psychological Review**, v. 93, p. 75-88, 1986.

HITCHCOCK, C. Contrastive Explanation and the Demons of Determinism. **British Journal for the Philosophy of Science**, v. 50, p. 585-612, 1999.

HITCHCOCK, C. The Intransitivity of Causation Revealed in Equations and Graphs. **Journal of Philosophy**, v. 98, p. 273-299, 2001.

HITCHCOCK, C. Do All and Only Causes Raise the Probabilities of Effects?. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 403-418.

HITCHCOCK, C. Prevention, Preemption, and the Principle of Sufficient Reason. **Philosophical Review**, v. 116, p. 495-532, 2007.

HITCHCOCK, C. Actual Causation: What's the Use?. *In:* BEEBEE, H.; HITCHCOCK, C.; PRICE, H. (ed.). **Making a Difference**: Essays on the Philosophy of Causation. Oxford: Oxford University Press, 2017, p. 116-31.

HITCHCOCK, C.; KNOBE, J. Cause and Norm. **Journal of Philosophy**, v. 106, p. 587-612, 2009.

HORWICH, P. Asymmetries in Time. Cambridge: MIT Press, 1987.

HUME, D. **An Enquiry concerning Human Understanding**, 1748.

ICARD, T.; KOMINSKY, J.; KNOBE, J. Normality and Actual Causal Strength. **Cognition**, v. 161, p. 80-93, 2017.

KAISERMAN, A. Causes and Counterparts. **Australasian Journal of Philosophy**, v. 95, p. 17-28, 2017.

KIM, J. Causation, Nomic Subsumption, and the Concept of Event. The Journal of Philosophy, v. 70, p. 217-36, 1973a

KIM, J. Causes and Counterfactuals. **The Journal of Philosophy**, v. 70, p. 570-572, 1973b.

KNOBE, J.; FRASER, B. Causal Judgment and Moral Judgment: Two Experiments. *In:* W. Sinnott-Armstrong (ed.). **Moral Psychology**, v. 2: The Cognitive Science of Morality: Intuition and Diversity. Cambridge: MIT Press, 2008, p. 441-448.

KOMINSKY, J.; PHILLIPS, J.; GERSTENBERG, T.; LAGNADO, D.; KNOBE, J. Causal Superseding. **Cognition,** v. 137, p. 196-209, 2015.

KUTACH, D. The Entropy Theory of Counterfactuals. **Philosophy of Science**, v. 69, p. 82-104, 2002.

KUTACH, D. **Causation and its Basis in Fundamental Physics**. Oxford: Oxford University Press, 2013.

KVART, I. Counterexamples to Lewis 'Causation as Influence'. **Australasian Journal of Philosophy**, v. 79, p. 411-423, 2001.

KVART, I. Causation: Probabilistic and Counterfactual Analyses. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 359-386.

LEWIS, D. **Counterfactuals**. Oxford: Blackwell, 1973a.

LEWIS, D. Causation. **Journal of Philosophy**, v. 70, p. 556-567, 1973b.

LEWIS, D. Counterfactual Dependence and Time's Arrow. **Noûs**, v. 13, p. 455-476, 1979.

LEWIS, D. A Subjectivist's Guide to Objective Chance. *In:* LEWIS, D. **Philosophical Papers**: v. II. Oxford: Oxford University Press, 1986a.

LEWIS, D. **Philosophical Papers**: v. II. Oxford: Oxford University Press, 1986a.

LEWIS, D. Events. 1986b. *In:* LEWIS, D. **Philosophical Papers**: v. II. Oxford: Oxford University Press, 1986a.

LEWIS, D. Postscripts to 'Causation'. 1986c. *In:* LEWIS, D. **Philosophical Papers**: v. II. Oxford: Oxford University Press, 1986a.

LEWIS, D. Causal Explanation. 1986d. *In:* LEWIS, D. **Philosophical Papers**: v. II. Oxford: Oxford University Press, 1986a.

LEWIS, D. **The Plurality of Worlds**. Oxford: Blackwell, 1986e.

LEWIS, D. Causation as Influence. **Journal of Philosophy**, v. 97, p. 182–97, 2000.

LEWIS, D. Causation as Influence (expanded version). 2004a. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 75-106.

LEWIS, D. Void and Object. 2004b. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 277-290.

LIPTON, P. Contrastive Explanation and Causal Triangulation. **Philosophy of Science**, v. 58, p. 687-697.

LIVENGOOD, J.; MACHERY, E. The Folk Probably Don't Think What You Think They Think: Experiments on Causation by Absence. **Midwest Studies in Philosophy**, v. XXXI, p. 107-128, 2007.

LOEWER, B. Counterfactuals and the Second Law. *In:* PRICE, H.; CORRY, R. (ed.). **Causation, Physics, and the Constitution of Reality**: Russell's Republic Revisited. Oxford: Oxford University Press, 2007, p. 293-326.

MASLEN, C. Causes, Contrasts, and the Nontransitivity of Causation. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 341-358.

MCDONNELL, N. Events and their Counterparts. **Philosophical Studies**, v. 173, p. 1291-1308, 2016.

MCDONNELL, N. Making a Contribution and Making a Difference. **American Philosophical Quarterly**, v. 55, p. 303-312, 2018.

MCGRATH, S. Causation by Omission. **Philosophical Studies**, v. 123, p. 125-148, 2005.

MELLOR, D. H. **The Facts of Causation**. London: Routledge, 1995.

MELLOR, D. H. For Facts as Causes and Effects. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 309-324.

MENZIES, P. Probabilistic Causation and the Pre-emption Problem. **Mind**, v. 105, p. 85-117, 1996.

MENZIES, P. Intrinsic versus Extrinsic Conceptions of Causation. *In:* SANKEY, H. (ed.). Causation and Laws of Nature. Kluwer: Academic Publishers, 1999, p. 313-329.

MENZIES, P. Difference-Making in Context. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 139-180.

MENZIES, P. Causation in Context. *In:* PRICE, H.; CORRY, R. (ed.). **Causation, Physics, and the Constitution of Reality**: Russell's Republic Revisited. Oxford: Oxford University Press, 2007, p. 191-223.

MENZIES, P. Platitudes and Counterexamples. *In:* BEEBEE, H.; HITCHCOCK, C.; MENZIES, P. (ed.). **The Oxford Handbook of Causation**. Oxford: Oxford University Press. 2009, p. 341-67.

MENZIES, P. The Problem of Counterfactual Isomorphs. *In:* BEEBEE, H.; HITCHCOCK, C.; PRICE, H. (ed.). **Making a Difference**: Essays on the Philosophy of Causation. Oxford: Oxford University Press, 2017, p. 153-74.

MONTMINY, M.; RUSSO, A. A Defense of Causal Invariantism. **Analytic Philosophy**, v. 57, p. 49-75, 2016.

NORTHCOTT, R. Causation and Contrast Classes. **Philosophical Studies**, v. 139, p. 111-123, 2008.

NOORDHOF, P. Probabilistic Causation, Preemption, and Counterfactuals. **Mind**, v. 108, p. 95-125, 1999.

NOORDHOF, P. 2004. Prospects for a Counterfactual Theory. *In:* DOWE, P.; NOORDHOF, P. (ed.). **Cause and Chance:** Causation in an Indeterministic World. London: Routledge, 2004, p. 188-201.

PAUL, L. A. Aspect Causation. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, 205-24.

PAUL, L. A.; HALL, N. **Causation:** A User's Guide. Oxford: Oxford University Press, 2013.

PEARL, J. **Causality**. Cambridge: Cambridge University Press, 2000.

PRICE, H. **Time's Arrow and Archimedes' Point**. Oxford: Oxford University Press, 1996.

PRICE, H.; CORRY, R. (ed.). **Causation, Physics, and the Constitution of Reality**: Russell's Republic Revisited. Oxford: Oxford University Press, 2007.

PRICE, H.; WESLAKE, B. The Time-Asymmetry of Causation. *In:* BEEBEE, H.; HITCHCOCK, C.; MENZIES, P. (ed.). **The Oxford Handbook of Causation**. Oxford: Oxford University Press, 2009, p. 414-443.

RAMACHANDRAN, M. A Counterfactual Analysis of Causation. **Mind**, v. 106, p. 263-277, 1997.

RAMACHANDRAN, M. A Counterfactual Analysis of Indeterministic Causation. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 387-402.

SCHAFFER, J. Trumping Preemption. **Journal of Philosophy**, v. 9, p. 165-181, 2000a.

SCHAFFER, J. Causation by Disconnection. **Philosophy of Science**, v. 67, p. 285-300, 2000b.

SCHAFFER, J. Contrastive Causation. **Philosophical Review**, v. 114, p. 297-328, 2005.

SCHULZ, K. If you'd wiggled A, then B would've changed: Causality and Counterfactual Conditionals. **Synthese**, v. 179, p. 239-251, 2011.

SPIRTES, P.; GLYMOUR, C.; SCHEINES, R. **Causation, Prediction, and Search**. 2. ed. New York: Springer, 2001.

STEGLICH-PETERSEN, J. Against the Contrastive Account of Singular Causation. **British Journal for the Philosophy of Science**, v. 63, p. 115-143, 2012.

WOODWARD, J. **Making Things Happen**: A Theory of Causal Explanation. Oxford: Oxford University Press, 2003.

WOODWARD, J. The Problem of Variable Choice. **Synthese**, v. 193, p. 1047-1072, 2016.

WOODWARD, J.; HITCHCOCK, C. Explanatory Generalizations. Part I: A Counterfactual Account. **Noûs**, v. 37, p. 1-24, 2003.

YABLO, S. De Facto Dependence. **Journal of Philosophy**, v. 99, p. 130-48, 2002.

YABLO, S. Advertisement for a Sketch of an Outline of a Prototheory of Causation. *In:* COLLINS, J.; HALL, N.; PAUL, L. (ed.). **Causation and Counterfactuals.** Cambridge: MIT Press, 2004, p. 119-138.

(V) As Leis da Natureza*

Autor: John W. Carroll
Tradução: Ariston Sócrates G. S. Cruz, Gabriel A. Eugênio & Kherian Gracher
Revisão: Karenina do Nascimento Rodrigues, Pedro Merlussi & Rhamon de Oliveira Nunes

Muitos princípios pertencentes ao domínio da ciência já foram, ou ainda são, considerados como leis da natureza: a lei da gravitação universal e as três leis do movimento de Newton, as leis dos gases ideais, as leis de Mendel, as leis da oferta e da procura, entre tantas outras. Outras regularidades importantes para a ciência não foram reconhecidas como dignas desse status; tais regularidades, contrariamente às leis, precisariam, ou ainda precisam, na visão de cientistas, de bases mais firmes; entre elas temos a regularidade das marés dos oceanos, o periélio da órbita de Mercúrio, o efeito fotoelétrico, a expansão do universo, etc. Cientistas usam as leis, em contraste a essas outras regularidades, para delimitar o que é possível: na cosmologia se reconhece a possibilidade do nosso universo

---

* CARROL, J. Laws of Nature. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2020. Disponível em: https://plato.stanford.edu/archives/win2020/entries/laws-of-nature/. Acesso em: 01 ago. 2021.

The following is the translation of the entry on Scientific Method, 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/win2020/entries/laws-of-nature/>. 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/archives/win2020/entries/laws-of-nature/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry.

ser aberto ou fechado porque tais possibilidades são consistentes com as leis gravitacionais de Einstein (MAUDLIN, 2007, p. 7-8). Na mecânica estatística, as leis de uma teoria física subjacente são utilizadas para determinar as trajetórias dinamicamente possíveis através do espaço de estados do sistema (ROBERTS, 2008, p. 12-16).

As filósofas e os filósofos da ciência e da metafísica discutem uma variedade de questões sobre as leis, mas a mais básica é esta: o que *é* uma lei? As duas respostas mais influentes incluem a teoria dos sistemas de David Lewis (1973, 1983, 1986, 1994), e a teoria de universais de David Armstrong (1973, 1983, 1986, 1994); outras abordagens incluem as posturas antirrealistas (VAN FRAASSEN, 1989; GIERE, 1999; WARD, 2002; MUMFORD, 2004) e as posturas antirreducionistas (CARROLL, 1994, 2008; LANGE, 2000, 2009; MAUDLIN, 2007). Além do problema mais básico, a literatura contemporânea também aborda questões sobre (i) se as leis sobrevêm aos fatos, (ii) o papel das leis no problema da indução, (iii) se envolvem necessidade metafísica, e (iv) o papel das leis na física e como isso se contrasta com o papel das leis nas ciências especiais.

1. A Questão Básica: O que é uma Lei?

Vejamos quatro motivos pelos quais filósofas e filósofos examinam a questão de saber o que é uma lei: em primeiro lugar, como dissemos acima, as leis parecem desempenhar um papel central na prática científica e, em segundo lugar, elas são importantes para muitos problemas filosóficos. Por exemplo, as filósofas instigadas e os filósofos instigados pela interpretação das contrafactuais defendida por Roderick Chisholm (1964,1965) e Nelson Goodman (1947), e também orientadas pelo modelo dedutivo nomológico de explicação de Carl Hempel e Paul Oppenheim (1948), levantaram a questão de saber o que faz com que as asserções contrafactuais e explicativas sejam verdadeiras; a convicção de que as leis devem ser relevantes para essa questão lhes motivou a questionarem o que distingue as leis daquilo que não é lei. A terceira razão é a sugestão bastante conhecida de Goodman de que há uma conexão entre **ser uma lei** e algo poder ser confirmado por meio de uma inferência indutiva. Isto posto, algumas pessoas adeptas da sugestão de Goodman acabaram por se deparar com o problema das leis em decorrência do interesse pelo problema da indução. Por fim, o quarto motivo é que filósofas e filósofos adoram um bom quebra-cabeça. Suponha que todas e todos aqui estejam sentados

(LANGFORD, 1941, p. 67). Então, trivialmente, é verdade que todas e todos estão aqui sentados. No entanto, apesar de verdadeira, tal generalização não parece ser uma lei, afinal, ela é demasiado acidental. O princípio de Einstein, por outro lado, de que nenhum sinal elétrico viaja mais rápido que a luz também é uma generalização verdadeira, mas é considerada uma lei; não é assim tão acidental. O que, então, diferencia esses casos?

Talvez o que expusemos acima não tenha a aparência de um quebra-cabeça. A asserção de que todo mundo está sentado aqui é espacialmente restrita, no sentido de que diz respeito a um lugar específico; o princípio da relatividade, por sua vez, não é restrito nesse sentido. Portanto é fácil pensar que, ao contrário das leis, as generalizações acidentalmente verdadeiras dizem respeito a lugares específicos. No entanto, isso não pode ser aquilo que as diferencia: existem não-leis que são verdadeiras que não são espacialmente restritas. Considere a generalização incondicional de que o diâmetro de todas as esferas de ouro é menor que 1 quilômetro. Não há esferas deste tamanho e o mais provável é que nunca haverá esferas assim; mas nem por isso diríamos que isso é uma lei. Também parece haver generalizações que poderiam expressar leis que são espacialmente restritas. A lei da queda livre de Galileu é a generalização de que, **na Terra**, os corpos em queda livre aceleram a uma taxa de 9.8 metros por segundo ao quadrado. Uma maneira mais clara de revelar a natureza intrigante desse quebra-cabeça é colocando a generalização das esferas de ouro lado a lado com uma generalização similar sobre as esferas de urânio:

> O diâmetro de todas as esferas de ouro é menor que 1 quilômetro
> O diâmetro de todas as esferas de urânio é menor que 1 quilômetro

Embora a primeira não seja uma lei, a última argumentavelmente o é, pois não é tão acidental quanto a primeira, dado que a massa crítica do urânio é tal que garante que uma esfera tão grande nunca existirá (VAN FRAASSEN, 1989, p. 27). Então qual é a diferença? O que faz com que a primeira seja uma generalização acidentalmente verdadeira e a última uma lei da natureza?

2. Os Sistemas

Uma resposta popular, a qual remonta a John Stuart Mill 1947, mas também foi defendida de uma forma ou outra por Frank Ramsey (1978), David Lewis (1973, 1983, 1986, 1994), John Earman (1984) e Barry Loewer (1996), estabelece uma conexão entre os sistemas dedutivos e aquilo que é uma lei. Os sistemas dedutivos são individuados pelos seus axiomas e as consequências lógicas dos axiomas são os teoremas. Alguns sistemas dedutivos verdadeiros serão mais fortes, outros serão mais simples. Essas duas virtudes, simplicidade e força, competem entre si (é fácil criar um sistema mais forte sacrificando a simplicidade: inclua todas as verdades como axiomas. Analogamente, é fácil criar um sistema simples sacrificando sua força: inclua apenas o axioma de que 2 + 2 = 4.). De acordo com Lewis (1973, p. 73), as leis da natureza pertencem a todos os sistemas dedutivos verdadeiros que possuem a **melhor** combinação entre simplicidade e força. Destarte, a ideia é, por exemplo, que a generalização de que o diâmetro de todas as esferas de urânio é menor que 1 quilômetro é uma lei porque é, discutivelmente, parte dos melhores sistemas dedutivos; a mecânica quântica é uma excelente teoria sobre nosso universo e talvez faça parte dos melhores sistemas, e é plausível pensar que a mecânica quântica somada a verdades sobre a natureza do urânio implicaria logicamente que não há esferas de urânio daquele tamanho (LOEWER, 1996, p. 112). Parece implausível dizer que a generalização do diâmetro de todas as esferas de ouro seja parte dos melhores sistemas; ela poderia ser adicionada como um axioma, mas não seria interessante em termos de força, e o seu acréscimo demandaria certos sacrifícios em termos de simplicidade. Lewis mais tarde revisou sua teoria significativamente para lidar com os problemas concernentes à probabilidade física (1986, 1994).

A teoria dos sistemas tem diversos aspectos atrativos. Em primeiro lugar, ela parece lidar bem com o problema da vacuidade das leis. Algumas leis são vacuamente verdadeiras: a primeira lei de Newton – todos os corpos inertes não possuem aceleração – é uma lei, apesar de não existirem corpos inertes. Mas também há muitas não-leis que são vacuamente verdadeiras: todo curupira pesa 5 quilos, todos os unicórnios são solteiros, etc. Com a teoria dos sistemas, as generalizações vacuamente verdadeiras não são excluídas do domínio das leis; mesmo assim apenas aquelas que pertencem aos melhores sistemas se qualificam como leis (LEWIS, 1986, p. 123). Ademais, é razoável pensar que um dos objetivos da teorização científica seja formular teorias verdadeiras que são bem balanceadas

em termos de simplicidade e força. Assim, a teoria dos sistemas parece sustentar o truísmo de que o objetivo da ciência é descobrir leis (EARMAN, 1978, p. 180; LOEWER, 1996, p. 112). Um último aspecto dessa teoria, considerado atrativo por muitas pessoas, embora não por todas, é que ela compactua com certas restrições da metafísica humiana sem fazer qualquer apelo explícito a conceitos modais (por exemplo, a condicional contrafactual, causalidade, disposições) e tampouco qualquer apelo evidente a entidades modalizantes (*vide* FOSTER, 2004). De fato, a teoria dos sistemas é a peça central da defesa de Lewis da sobreveniência humiana, "a doutrina de que tudo o que existe no mundo é um vasto mosaico de fatos locais e particulares, apenas uma coisinha aqui e depois outra acolá." (1986, IX).

Por outro lado, filósofas e filósofos ficam, respectivamente, receosas e receosos em relação a outros aspectos da teoria dos sistemas. Confira, em particular, Armstrong (1983, p. 66-73), van Fraassen (1989, p. 40-64) e Carroll (1990, p. 197-206). Algumas pessoas argumentam que a teoria tem a consequência desagradável de que as leis são inapropriadamente dependentes da nossa mente, em razão de seu apelo a conceitos de simplicidade, força e melhor balanceamento, que são conceitos cuja instanciação parece depender de habilidades cognitivas, interesses e propósitos. O apelo à simplicidade levanta questões adicionais provenientes da aparente necessidade de uma linguagem regimentada para permitir comparações razoáveis entre sistemas (LEWIS, 1983, p. 367). Mais contemporaneamente, John Roberts (2008, p.10) questionou a teoria dos sistemas num ponto que fora outrora considerado uma virtude da teoria: “Não temos qualquer prática em pesar virtudes competidoras de simplicidade e conteúdo informacional com o propósito de escolher um sistema dedutivo em detrimento de outros, nos quais são todos presumivelmente verdadeiros”. Há a prática de ajuste de curvas, que envolve pesar as virtudes competidoras de simplicidade e qualidade de ajuste, mas esta última não é uma prática que é parte do processo de descobrimento **do que é verdadeiro**. Tim Maudlin (2017, p. 16) e Roberts (2008, p.23) também objetam que a teoria dos sistemas não é adequada para descartar aquelas regularidades evidentes, encontradas em toda parte como leis. Por exemplo, regularidades expressas por “o universo é fechado”, “a entropia geralmente cresce”, “os planetas do nosso sistema solar são coplanares”, entre outras, se verdadeiras, poderiam ser adicionadas a qualquer sistema dedutivo verdadeiro, aumentando consideravelmente sua força, com apenas um pequeno custo em termos de simplicidade. Interessante notar que, às vezes, a teoria dos sistemas é abandonada **justamente** por satisfazer as restrições humianas gerais em uma concepção de leis da natureza; algumas e alguns argumentam que

o fato de algumas generalizações serem leis não é determinado por questões locais de fatos particulares (*vide* seção 4). Embora humianos como Lewis geralmente prefiram o realismo em detrimento de qualquer forma de antirrealismo (*vide* seção 5), Nora Berenstain e James Ladyman (2012) já defenderam que o realismo científico é incompatível com o humianismo porque o primeiro requer uma noção de necessidade natural, a qual não é suscetível a um tratamento humiano.

3. Universais

Uma nova concorrente à teoria dos melhores sistemas, e também a outras teorias humianas das leis da natureza, surgiu no final dos anos 70. Encabeçada por David Armstrong (1978, 1983, 1991, 1993), Fred Dretske (1977), e Michael Tooley (1977, 1987), tal concepção rival apela a universais para distinguir as leis daquilo não é uma lei.

Foquemos no tratamento que Armstrong ofereceu desta visão; vejamos uma de suas formulações mais concisas da visão geral característica dessa concepção a partir dos universais:

> Suponha que seja uma lei que todos *Fs* são *Gs*. Tanto *F* quanto *G* são tidos como universais. Certa relação, uma relação de necessitação contingente, ou não-lógica, se dá entre *ser F* e *ser G*. Este estado de coisas pode ser simbolizado como '*N(F,G)*'. (1983, p. 85).

A perspectiva acima promete resolver alguns enigmas e problemas familiares: talvez a diferença entre a generalização das esferas de urânio e a das esferas de ouro seja esta, a saber, que ser urânio de fato necessita ter menos de um quilômetro de diâmetro, ao passo que o mesmo não pode ser dito a respeito do ouro. Aqui não aparecem as preocupações sobre a natureza subjetiva da simplicidade, força e melhor equilíbrio; não há a ameaça de que ser uma lei possa ser dependente da mente humana. Alguns e algumas (ARMSTRONG, 1991; DRETSKE, 1977) pensam que o quadro geral da concepção dos universais dá suporte à ideia de que as leis desempenham um papel explicativo especial nas inferências indutivas, uma vez que uma lei não é apenas uma generalização universal verdadeira, mas uma coisa completamente diferente, uma relação existente entre dois universais. Essa

perspectiva também é consistente com a ideia de que ser uma lei da natureza não sobrevém a questões de fatos particulares; a negação da sobreveniência humiana é amiúde acompanhada da aceitação da concepção de universais.

No entanto, para que realmente ocorra este desfecho, mais precisa ser dito sobre **N**. Este é o problema que Bas van Fraassen denominou como o **problema da identificação**, o qual se junta com um segundo problema, a saber, o **problema da inferência** (1989, p. 96). A essência desse par de problemas foi percebida logo de início por Lewis, com seu faro usual:

> Seja lá o que *N* for, não consigo ver como poderia ser absolutamente impossível ter *N (F, G)* e *Fa* sem *Ga* (a não ser que *N* seja apenas uma conjunção constante, ou uma conjunção constante somada a alguma outra coisa, em tal caso a teoria do Armstrong se transformando num tipo de teoria de regularidade que ele rejeita). Esse mistério é meio que escondido pela terminologia de Armstrong. Ele usa 'necessita' para o universal legislador *N*; e quem se surpreenderia em ouvir que se *F* necessita *G* e *a* tem *F*, então *a* deve ter *G*? Mas eu diria que *N* merece o nome 'necessitação' somente se, de algum modo, ele realmente puder entrar nas conexões necessárias requeridas. Não pode entrar nelas só por possuir um nome, do mesmo modo como alguém não pode ter grandiosos bíceps[15] só por ser chamado de 'Armstrong'. (1983, p. 366).

Basicamente, é preciso que haja uma especificação acerca do que é a relação legisladora (o problema da identificação). Destarte é preciso determinar se tal relação é adequada para a tarefa (o problema da inferência): A relação *N* se obter entre *F* e *G* implica que *F*s são *G*s? Esta relação suporta contrafactuais correspondentes? As leis realmente não sobrevêm, são independentes da mente humana, são explicativas? Armstrong de fato diz mais sobre o que é a relação legisladora *N*. Em resposta a Fraassen, o autor afirma:

---

[15] N.T.: "braços fortes".

> É a essa altura que, afirmo, o problema da identificação foi resolvido. A relação requerida é uma relação causal, (…) agora conjecturada como relacionando *tipos* e não *espécimes*. (1993, p. 422).

Ainda restam perguntas sobre a natureza dessa relação causal entendida como a relação que relaciona tanto espécimes de eventos quanto universais (ver van Fraassen 1993, p. 435–437, e Carroll 1994, 170–174).

4. A Sobreveniência humiana

Em vez de tentar detalhar todos aqueles elementos críticos que separam as teorias dos sistemas e de universais, será mais profícuo se dirigirmos nosso foco à questão da sobreveniência (a determinação), a qual é particularmente divisora. Trata-se aqui de saber se as considerações humianas realmente determinam o que são as leis. Existem exemplos importantes que parecem mostrar que a resposta seja negativa.

Tooley (1977, p. 699) nos pede para supor que há dez tipos diferentes de partículas fundamentais, portanto se segue que há cinquenta e cinco tipos diferentes de interação entre duas partículas. Suponha que cinquenta e quatro tipos dessas interações já tenham sido estudadas, resultando na descoberta de cinquenta e quatro leis. Mas suponha que a interação entre as partículas de tipo *X* e *Y* ainda não foram estudadas porque as condições são tais que elas nunca irão interagir. De qualquer modo, parece que pode existir uma lei de que quando as partículas *X* e as partículas *Y* se interagem, *P* ocorre. Analogamente, talvez possa ser uma lei que sempre que as partículas *X* e *Y* interagem, *Q* ocorre. Parece não haver nada sobre os fatos particulares e locais deste mundo que fixe qual dessas duas generalizações seja uma lei.

O fracasso da sobreveniência sugerido pelo exemplo de Tooley também surge em outros casos. Considere a possibilidade de existir uma partícula solitária viajando na velocidade de, digamos, um metro por segundo através de um espaço que, sem ela, seria vazio. Parece inclusive que esse universo poderia ser um universo Newtoniano quase vazio no qual é acidentalmente verdadeiro que todos os corpos têm velocidade de um metro por segundo; ocorre apenas que nada mais existe naquele universo para alterar o movimento da partícula. No entanto, poderia

também ser o caso que esse mundo não seja Newtoniano e que seja uma lei que todos os corpos têm velocidade de um metro por segundo; poderia ser o caso que essa generalização não fosse acidental e continuaria verdadeira mesmo se houvesse outros corpos colidindo com a partícula solitária (ver especialmente Earman 1986, p. 100; Lange 2000, p. 85-90).

Maudlin avança uma objeção aos humianos baseada na prática comum entre as físicas e os físicos ao considerar modelos das leis de uma teoria:

> O espaço de Minkowski, o espaço-tempo da teoria especial da Relatividade, é um modelo das equações de campo da Relatividade Geral (em particular, é uma solução vazia). Então um espaço de Minkowski vazio é um jeito que o mundo poderia ser caso fosse governado pelas leis da Relatividade Geral. Mas o espaço de Minkowski é *exclusivamente* um modelo para leis de Relatividade Geral? Claro que não! Alguém poderia, por exemplo, postular que a Relatividade Especial é a descrição completa e acurada da estrutura do espaço tempo, e produzir outra teoria gravitacional, e que poderia ainda ter o espaço de Minkowski vazio como modelo. Então, assumindo que nenhum mundo possível poderia ser governado ao mesmo tempo pelas leis da Relatividade Geral e pelas leis de uma teoria gravitacional rival, teríamos que o espaço físico total do mundo não pode sempre determinar as leis (2007, p. 67).

Aqui temos a sugestão de que existe a possibilidade de um universo sem matéria com as leis da Relatividade Geral e outro com as leis de uma teoria gravitacional opositora (para mais exemplos, ver Carroll 1994, p. 60-80). O que Maudlin vê como consequência do raciocínio científico padrão, humianos enxergam como um exemplo do absurdo da não-sobreveniência.

Humianos avançam a tese de que vários dos ditos mundos possíveis não são de fato possíveis. Por vezes essa tese conta com a dificuldade sobre governança das leis, ou às vezes preocupações de natureza epistemológica ou ontológicas, ou ainda dificuldades sobre o funcionamento da linguagem. Uma objeção humiana à não-sobreveniência vem de Helen Beebee (2000). O ponto dela é que se alguém

se inserir no debate com a concepção de governança em mente vai considerar os exemplos anti-sobreveniência convincentes, mas que apelar a essa concepção para rejeitar de alguma maneira a análise humiana consiste em uma petição de princípio ou em um ponto não convincente, já que esta é uma concepção rejeitada por humianos (ver também Loewer 1996 e Roberts 1998). Em contraste, Susan Schneider (2007), Barry Ward 2007, e Roberts (2008) simpatizam com aspectos do humianismo e aspectos da concepção de governança.

Em dois artigos, Earman e Roberts (2005a e b) inicialmente discutem qual seria a melhor formulação da tese da sobreveniência humiana, e então argumentam, com base em considerações céticas, a favor de sua própria versão da sobreveniência humiana. Jonathan Schaffer (2008, p. 96–97, p. 94-99) rejeita as considerações céticas (também ver Carroll 2008, p. 75–79), mas avança a preocupação ontológica de que as leis não-sobrevenientes são entidades infundadas (84-85).

Roberts (2008, 357-61) oferece uma forma original de responder aos aparentes contraexemplos à sobreveniência. No exemplo descrito acima da partícula solitária, existe um mundo em que a partícula solitária está viajando a um metro por segundo, embora não seja uma lei que todas as partículas viajam nessa velocidade. Há também um mundo em que a partícula solitária viaja a um metro por segundo, embora seja uma lei que todas as partículas estejam viajando nessa velocidade. Segundo Roberts, este raciocínio não é contraditório com a sobreveniência devido à sensitividade ao contexto do predicado "é uma lei". Embora a sentença "É uma lei que todas as partículas viajam a velocidade de um metro por segundo" seja (i) verdadeira em relação a um par contexto/mundo e (ii) falsa em relação a outro par contexto/mundo, essa diferença no valor de verdade poderia simplesmente ser o resultado da diferença entre dois contextos.

Sem entrar em muitos detalhes sobre sua visão metateórica sobre o que é ser uma lei, Roberts defende que para um mundo possível *m* em que existe somente uma partícula solitária viajando em velocidade constante durante toda história, e relativa ao contexto em que a teoria predominante é, digamos, a mecânica Newtoniana, a sentença "É uma lei que todas partículas viajam a velocidade de um metro por segundo" é verdadeira somente se a referência da cláusula "que" figura o papel de lei na teoria predominante, o que não acontece neste caso. Ela poderia desempenhar o papel de lei em relação à outra teoria, mas isso estaria vinculado a um contexto diferente, uma vez que a teoria predominante teria de ser diferente. De acordo com Roberts, uma única generalização não pode ao mesmo tempo ter e não ter o papel de lei em relação a uma teoria e, portanto, tanto uma teoria predominante

diferente quanto um contexto diferente seriam requeridos para que "É uma lei que todas as partículas viajam a velocidade de um metro por segundo" seja verdadeira.

O que torna a resposta acima atraente é que ela não rejeita qualquer afirmação intuitiva sobre as leis nos diversos mundos possíveis (Roberts 2008, p. 360); os juízos anti-sobreveniência sobre o que são leis passam a ser afirmações razoáveis dado o contexto; a única coisa é que fracassam em reconhecer a influência do mesmo. Então, por exemplo, as supostas duas possibilidades de Maudlin seriam interpretadas por Roberts como descrições de uma mesma possibilidade feitas em relação a dois contextos com teorias predominantes diferentes: a teoria Geral da Relatividade e alguma teoria gravitacional rival (pontos paralelos seriam feitos a respeito dos exemplos de Tooley envolvendo 10 tipos diferentes de partículas fundamentais). Aqui, a chave é a sensitividade ao contexto que Robert imputa nas condições de verdade de sentenças sobre *ser uma lei*. Ainda há outras visões que interpretam o que é *ser uma lei* como sensíveis ao contexto e podem se aproveitar do modo que Roberts desafia os exemplos anti-sobreveniência. No entanto, o que não é persuasivo na posição de Roberts é sua concepção sobre as atribuições de legalidade das leis serem dependentes do contexto. Isto porque sua perspectiva é desenvolvida para uma expressão particular do inglês: "*law of nature*", mas seria bem melhor se o tratamento contextual de "lei da natureza" combinasse com a dependência ao contexto de palavras e expressões de outras linguagens naturais. Deveríamos tentar entender nosso uso de "lei da natureza" como dependente ao contexto por meio do nosso apelo a princípios linguísticos, e a investigação deveria ser motivada pelas práticas conversacionais (Carrol 2018, 131-32). O termo inglês "*law of nature*" não deveria ser uma anomalia isolada da língua inglesa (cf., Unger 1971, 202) sobre o verbo "conhecer".

5. Antirrealismo

A maioria das filósofas contemporâneas e dos filósofos contemporâneos é *realista* sobre as leis; acreditam que alguns relatos sobre o que são as leis são bem-sucedidos em descrever a realidade. No entanto, existem algumas e alguns *antirrealistas* que discordam disso.

Por exemplo, van Fraassen, Ronald Giere, e também Stephen Mumford, acreditam que não existem leis. Van Fraassen encontra suporte à sua visão nos

problemas que afligem as abordagens de Lewis e de Armstrong, e o considerado fracasso de Armstrong e outras/outros em prover uma descrição de uma epistemologia que permita a crença racional em leis (1989, p. 130 e p. 180-181). Giere apela para as origens do uso do conceito de lei na história da ciência (1999, p.86-90) e defende que as generalizações frequentemente descritas como leis não são de fato verdadeiras (90-91). As razões de Mumford são mais metafísicas; ele sustenta que, para governar, leis devem ser externas às propriedades que elas governam, mas, para serem externas de tal modo, as propriedades governadas devem carecer de condições de identidade adequadas (2004, p. 144-145). Outras e outros adotam uma variedade sutilmente diferente de antirrealismo. Embora essas pessoas possam proferir sentenças como 'É uma lei que nenhum sinal elétrico viaja mais rápido que a luz', elas são antirrealistas em virtude de pensar que tais sentenças não são (puramente) asserções de fatos. Se esta generalização Einsteiniana é ou não uma lei não é um fato sobre o universo; não é algo lá fora à espera de sua descoberta. Relatos sobre o que são leis apenas projetam certa atitude (em adição à crença) sobre as generalizações contidas neles. Por exemplo, Ward (2002, p. 197) toma atitude como sendo uma acerca da adequação da generalização para predição e explicação (também ver Blackburn 1984 e 1986).

O desafio para o antirrealismo é minimizar o caos que uma realidade sem leis poderia se impor em nossas práticas de senso comum e científicas. Acerca da ciência, os exemplos e usos de leis descritos no início deste verbete atestam que 'lei' possui um valor perceptível na ciência e que cientistas parecem preparadas e preparados para tomar como factivo. Sobre nossas práticas de senso comum, embora 'lei' não seja frequentemente parte de nossas conversas ordinárias, um antirrealismo sobre leis ainda poderia ter consequências abrangentes. Isto se dá porque o conceito de ser uma lei está relacionado a outros conceitos, especialmente aqueles nômicos, conceitos como o condicional contrafactual, disposições e causação. Por exemplo, parece que deve haver pelo menos uma lei da natureza para haver quaisquer verdades contrafactuais interessantes. Se um fósforo ordinário em condições usuais fosse riscado, ele acenderia? Parece que sim, mas somente porque nós presumimos que a natureza é regular de certos modos. Nós pensamos que este contrafactual é verdadeiro porque acreditamos que há leis. Não houvesse leis, não seria o caso que um fósforo seria aceso se fosse riscado. Como resultado, teríamos que também não seria o caso que o fósforo tinha a *disposição* de acender, ou que riscá-lo *causaria* que ele fosse aceso.

Uma ou um antirrealista poderia se esquivar desse desafio negando as conexões entre o ser uma lei e outros conceitos? Isto permitiria a uma pessoa ser antirrealista sobre as leis, mas continuar sendo realista sobre, digamos, as contrafactuais? Aqui, o perigo à espreita é que a postura resultante parece fadada a ser *ad hoc*. Conceitos tais como a condicional contrafactual, disposições, e causação exibem muitos dos mesmos aspectos intrigantes das leis; parece haver problemas e *puzzles* filosóficos paralelos com respeito a esses conceitos. É difícil ver o que suportaria o antirrealismo sobre as leis da natureza, mas não também sobre outros conceitos nômicos.

6. Antirreducionismo

John Carroll (1994, 2008), Marc Lange (2000, 2009), e Maudlin (2007) advogam por visões antirreducionistas e anti-sobreveniência (ver também Ismael 2015 e Woodward 1992). Com respeito a questões sobre o que é uma lei, todos rejeitam as respostas dadas por humianos como Lewis, negam a sobreveniência humiana e não veem vantagem em apelar para universais. Eles rejeitam todas as tentativas de dizer o que é uma lei que não apelam a conceitos nômicos. Porém, eles ainda acreditam que há, com efeito, leis da natureza, ou seja, não são antirrealistas.

Maudlin (2007, p. 17-18) entende a legalidade da lei como possuidora de um *status* primitivo, e as leis como sendo primitivos ontológicos — entidades fundamentais em nossa ontologia. Deste modo, seu projeto consiste em mostrar quais operações as leis podem fazer, definindo a possibilidade física em termos das leis e esboçando uma interpretação das condicionais contrafactuais e da explicação baseada em leis.

Carroll (2008) esboça uma análise do que faz algo ser uma lei em termos de conceitos causais/explanatórios. O ponto inicial é a intuição de que as leis não são acidentais, que elas não são coincidências. Não ser uma coincidência, entretanto, não é tudo que há para algo não ser uma lei. Por exemplo, pode ser verdade que não haja nenhuma esfera de ouro maior que 1000 quilômetros em diâmetro porque há somente peças de ouro muito pequenas no universo. Neste caso, estritamente falando, essa generalização seria verdadeira, de maneira adequadamente geral, e não seria coincidência. Não obstante, essa não poderia ser uma lei. Discutivelmente, o que bloqueia tal generalização de ser uma lei é que algo *na* natureza — na verdade, uma condição inicial do universo, a quantidade limitada de ouro — é responsável

pela generalização. Contraste isso com a lei que diz que os corpos inerciais não possuem aceleração. Com essa e outras leis, parece que elas se seguem da (própria) natureza.

O tratamento de Lange (2000, 2009) inclui uma interpretação do que é ser uma lei em termos de uma noção contrafactual de estabilidade. A explicação geral é complexa, mas a ideia básica é esta: seja um conjunto logicamente fechado de proposições estável se, e somente se, os membros do conjunto permaneceriam verdadeiros dado qualquer antecedente que seja consistente com o próprio conjunto. Assim, por exemplo, o conjunto de verdades lógicas é trivialmente estável, já que verdades lógicas seriam verdadeiras sempre. Um conjunto que inclui uma generalização acidental de que todas as pessoas na sala estão sentadas, mas é consistente com a proposição que alguém da sala grita "Fogo!", não seria um conjunto estável; se alguém gritasse "Fogo!", então alguém na sala não estaria sentado. Lange argumenta (2009, p. 34) que nenhum conjunto estável de fatos sub-nômicos — exceto, talvez, o conjunto de todas as verdades — contém alguma verdade acidental. "Por identificar as leis como os membros de ao menos um conjunto estável não maximal, nós descobrimos como a legalidade de um fato sub-nômico é fixada pelos fatos sub-nômicos, e pelos fatos subjuntivos sobre eles" (2009, p. 43).

As tentativas de enfraquecer o antirreducionismo frequentemente incluem certos desafios para a anti-sobreveniência, como aquelas mencionadas ao final da Seção 4. Tyler Hildebrand (2013) desafia o antirreducionismo de Carrol e Maudlin baseado no fato das leis primitivas não serem bem-sucedidas em explicar a uniformidade da natureza. Um simpósio sobre o livro *Laws and Lawmakers* (2009) inclui, entre as respostas de Lange, uma variedade de críticas de Carroll, Loewer e James Woodward (ver Lange *et al*., 2011). Heather Demerest (2012) levanta três objeções para o antirreducionismo de Lange, todas centradas na questão de saber se as condicionais no subjuntivo são adequadas para exercer a função legisladora de fazer as leis.

7. Indução

Goodman pensou que a diferença entre as leis da natureza e as verdades acidentais estaria ligada intrinsecamente ao problema da indução. Em seu *The new Riddle of Induction* (1983, [1ª edição 1954], p. 73), Goodman diz:

> Apenas uma afirmação que é *legiforme* – independentemente de sua verdade ou falsidade, ou de sua importância científica – é capaz de ser confirmada a partir de uma de suas instâncias. Este não é o caso de afirmações acidentais.

(Terminologia: *P* é *legiforme* apenas se *P* é uma lei verdadeira). Goodman reivindica que, se uma generalização é acidental (não *legiforme*), então ela não é capaz de receber confirmação a partir de alguma de suas instâncias.

Isso provocou muita discussão, incluindo algumas objeções. Por exemplo, suponha dez lançamentos de uma moeda comum. Os primeiros lançamentos deram cara (Dretske 1977, p. 256-257), as nove primeiras instâncias — pelo menos em um sentido — confirmam a generalização de que todos os lançamentos darão cara. A probabilidade dessa generalização seria (0,5)10 até 0,5. Porém, essa generalização não é *legiforme*, pois se fosse verdadeira não seria uma lei. A resposta padrão a esse exemplo pode ser através do argumento de que essa não é a noção pertinente de confirmação (sendo apenas um "recorte de conteúdo"), sugerindo que o *legiforme* requer a confirmação de instâncias não examinadas da generalização. Note que, no caso da moeda, a probabilidade de que a décima moeda cairá cara não altera o fato de ter caído cara depois dos primeiros nove lançamentos. Há, no entanto, exemplos que geram problemas para essa ideia também.

> Imagine que a gente tenha um quarto com cem homens, e você pergunte para 50 se eles são terceiros filhos. Eles respondem que são. Seria razoável que, de alguma forma, ao menos segundo as nossas expectativas, o próximo que você fosse perguntar também respondesse que é um terceiro filho (Jackson e Pargetter 1980, p. 423).

Não seria bom revisar a afirmação para dizer que nenhuma generalização *que se acredita* ser acidental é capaz de confirmação. Sobre o caso do terceiro filho, alguém poderia saber que a generalização, ainda que verdadeira, não é uma lei. E a discussão continua. Frank Jackson e Robert Pargetter propuseram uma conexão alternativa entre confirmação e leis, na qual certas verdades contrafactuais deveriam se seguir: a observação de que tudo aquilo que é *A* também é *F* e *B* confirma que todos aqueles que são *A* e não-*F* são Bs apenas se aquilo que é *A*

teria sido tanto *A* quanto *B* se não tivesse sido *F* (essa sugestão foi criticada por Elliott Sober 1988, p. 97-98). Lange (2000, p. 111-142) usa uma estratégia diferente. Ele tenta refinar ainda mais a noção relevante de confirmação, caracterizando uma noção intuitiva da confirmação indutiva. Assim, ele mantém que as únicas generalizações que não são consideradas como não-*legiformes* podem ser confirmadas indutivamente.

Às vezes a ideia de que as leis possuem um papel especial a desempenhar na indução serve como ponto de partida para críticas às análises humianas. Dretske (1977, p. 261-262) e Armstrong (1983, p. 52-59, e 1991) adotam um modelo de inferências indutivas que envolve uma inferência a favor da melhor explicação (ver também Foster 1983 e 2004). Nesse constructo simples, o modelo descreve um padrão que começa com uma observação de instâncias de uma generalização, inclui uma inferência para a lei correspondente (essa é a inferência para a melhor explicação), e conclui com uma inferência para a própria generalização ou para suas instâncias não observadas. A alegação posta contra as/os humianos é que, em sua visão sobre o que são leis, leis não são adequadas para explicar suas instâncias, então não conseguem sustentar a requerida inferência para a melhor explicação.

Essa é uma área onde o trabalho sobre as leis precisa ser feito. Armstrong e Dretske fazem afirmações substantivas sobre o que pode e o que não pode ser uma instância confirmada: *grosso modo*, as leis humanas não podem, enquanto leis, como universais, podem. Mas, pelo menos, essas afirmações não podem estar inteiramente corretas. As leis humianas não podem ser uma instância confirmada? Como a discussão acima ilustra, Sober, Lange e outras/outros argumentaram que as generalizações conhecidas por serem acidentais podem ser confirmadas por suas instâncias. Dretske e Armstrong precisam de uma premissa plausível e apropriadamente forte conectando o que faz algo ser uma lei com confirmabilidade, mas não é claro que haja tal premissa para ser obtida. Vejamos o problema básico: Como muitas autoras e muitos autores têm notado (e.g., Sober 1988,98; van Fraassen 1987, 255), a confirmação de uma hipótese ou suas instâncias não examinadas sempre serão sensíveis a crenças de fundo presentes. Tanto que, com um fundo de crenças do tipo certo, praticamente tudo pode ser confirmado, independente de seu *status* de lei ou de sua *legiformidade*. Assim, afirmar um princípio plausível descrevendo a conexão entre as leis e o problema da indução seria difícil.

8. Necessidade

A filosofia geralmente considera que algumas verdades contingentes são (ou poderiam ser) leis da natureza. Além disso, pensa-se que, se há uma lei que todos *F*s são *G*s, então não é preciso haver qualquer conexão (metafisicamente) necessária entre *ser F e ser G*, ou seja, é (metafisicamente) possível que algo seja *F* e não seja *G*. Por exemplo, algum mundo possível que, em questão de leis, obedeça ao princípio *geral* da física Newtoniana, é um mundo em que a primeira lei de Newton é verdadeira, e um mundo contendo aceleração de corpos inerciais é um mundo em que a primeira lei de Newton é falsa. Este último é também o mundo no qual a inércia é instanciada, mas não necessita aceleração zero. Necessitarianistas, no entanto, asseguram que todas as leis são verdades necessárias (ver Shoemaker 1980 e 1998, Swoyer 1982, Fales 1990, Bird 2005. Ver Vetter 2012 pela crítica ao Bird 2005 no campo do essencialismo disposicional). Tem-se defendido algo que é ligeiramente diferente. Mantendo que algumas leis são afirmações singulares sobre universais, permite-se que algumas leis são contingentemente verdadeiras. Então, nessa visão, uma lei como F/G pode ser falsa se F não existir; mesmo assim essa diferença é menor. Certas autoras e certos autores pensam que para existir uma lei F/G deve ser uma verdade necessária que todos os Fs são Gs (ver Tweedale 1984, Bigelow, Ellis, e Lierse 1992, Ellis e Lierse 1994, e Ellis 2001, p. 203-228; 2009, p. 57-72).

Duas razões podem ser dadas para acreditar que uma lei não depende de alguma conexão necessária entre propriedades. A primeira razão é a concebilidade de que em um mundo possível uma lei seja todos *F*s são *G*s apesar de haver outro mundo com um *F* que não é *G*. A segunda é que há leis que podem apenas serem descobertas de uma maneira *a posteriori*. Se a necessidade é sempre associada com as leis da natureza, então não é claro por que as/os cientistas não podem sempre proceder através de métodos *a priori*. Naturalmente, essas duas razões são, com frequência, desafiadas. As e os necessitarianistas argumentam que concebilidade não é um guia para a possibilidade. Elas e eles também apelam ao argumento de Saul Kripke (1972), importante por revelar certas verdades *a posteriori*, com o intuito de argumentar que a natureza *a posteriori* de algumas leis não previne aquilo que as fazem serem leis de requerer uma conexão necessária entre as propriedades. Em suporte adicional para sua própria visão, as e os necessitarianistas argumentam que sua posição é uma consequência da teoria de disposições que adotam, de acordo com a qual disposições possuem seu poder causal essencialmente.

Então, por exemplo, nessa teoria, a carga tem como parte de sua essência o poder de repelir cargas semelhantes. As leis, então, são acarretadas pelas essências das disposições (BIRD, 2005, p. 356). Do modo como necessitarianistas veem, é uma virtude dessa posição poder explicar o porquê de as leis suportarem contrafactuais; elas suportam contrafactuais da mesma maneira que outras verdades necessárias o fazem (SWOYER, 1982, p. 209; FALES, 1990, p. 85-87).

A preocupação primária para as e os necessitarianistas concerne a sua habilidade de sustentar suas rejeições das razões tradicionais para pensar que algumas leis são contingentes. O problema (SIDELLE, 2002, p. 311) é que elas e eles também fazem distinções entre verdades necessárias e contingentes, e parecem até mesmo confiar em considerações de concebilidade para fazê-lo. À primeira vista, não há nada de especialmente suspeito sobre o julgamento de que é possível que um objeto viaje mais rápido do que a luz. Como isso pode ser pior do que o julgamento de que é possível que esteja chovendo em Paris? Outro problema para as e os necessitarianistas é se seu essencialismo sobre disposições pode sustentar todos as contrafactuais que são aparentemente suportados pelas leis da natureza (LANGE, 2004).

9. Leis, Circularidade e as Perspectivas de Explicação

A teoria humiana das leis tem sido desafiada de diferentes maneiras pelo menos desde Armstrong (1983, p. 40), e um desses desafios consiste no problema de saber se as leis humianas são, de fato, explicativas. Este último é, mais contemporaneamente, formulado de maneira perspicaz por Maudlin:

> Se se é humiano, então parece que o próprio Mosaico humiano não admite qualquer explicação adicional. Visto que se trata do alicerce ontológico em termos do qual todas as outras coisas existentes devem ser explicadas, nenhuma dessas coisas pode realmente explicar a estrutura do próprio mosaico. Esta crítica inclusive se reverbera há anos como uma objeção comum a qualquer teoria humiana das leis da natureza. Se as leis não são nada além das características genéricas do mosaico humiano, então há um sentido no qual não se pode apelar a essas próprias leis no

> intuito de explicar as características particulares do mosaico: se as leis são o que são, então isso de dá em virtude do Mosaico e não vice-versa (MAUDLIN, 2007, p. 172).

Lower (2012, p. 131) oferece uma resposta à questão acima destacada por Maudlin. A resposta de Loewer é que o grande mosaico humiano é o faz as leis da natureza serem verdadeiras. Para evitar a objeção de circularidade, ele diz que as leis humianas não explicam **metafisicamente** os elementos do mosaico, mas elas o fazem **cientificamente**, sugerindo que há duas noções de explicação e, portanto, nenhuma circularidade.

Uma forma cada vez mais popular de entender a relação entre as leis e suas instâncias consiste em considerar tais instâncias como o fundamento das leis da natureza. Embora nenhuma instância particular de uma lei da natureza possa fundamentá-la completamente, uma conjunção de tais instâncias pode. Outra maneira plausível de entender a relação entre as leis e suas instâncias é entender a lei como aquilo que fundamenta suas instâncias (EMERY, 2019). Porém, como a relação de fundamentação é assimétrica, as duas visões acima descritas não podem ser ambas verdadeiras. Há uma saída para esse dilema que é reveladora para o presente debate. Note que embora (P & Q) seja uma fundamentação completa de Q, parece errado afirmar que (P & Q) explica o porquê de *Q* (ROSKI, 2018). Isso acontece por causa do fato do conteúdo do *explanandum* (aquilo que é para ser explicado) vir acoplado ao conteúdo do *explanans* (aquele cujo objetivo é explicar), e nada pode explicar a si próprio (ou ser uma parte essencial da explicação de si próprio). Veja que essa formulação expõe o problema: se o *explanans* inclui o *explanandum* como parte de seu conteúdo, então ele torna a explicação desnecessária, pois já deveríamos ter de antemão algum entendimento do *explanandum.* Como as explicações bem-sucedidas não são circulares, qualquer pessoa que aceite que as leis fundamentem suas instâncias não deve pensar que a relação de fundamentação seja explicativa. Vale notar que isso não quer dizer que a relação de fundamentação não seja explicativa, mas sim que as **leis da natureza** não são adequadas para explicar suas instâncias. A circularidade também afeta o modelo dedutivo-nomológico de explicação, como as autoras e os autores de tal modelo salientaram:

> [...] O conteúdo do *explanandum* está contido no explanans. Isso está correto porque o explanandum

> é uma consequência semântica do explanans (*vide* HEMPEL; OPPENHEIM, 1948, p. 162; SCHUMENER 2017, p. 793).

O que se questiona aqui é a importância do papel que as explicações têm em oferecer entendimento. A validade requerida trás consigo circularidade semântica porque o conteúdo do *explanans* seria suficiente para a verdade do *explanandum;* as formulações típicas do modelo dedutivo-nomológico requerem ao menos uma lei da natureza como premissa de um "argumento explicativo". Com efeito, ao menos uma lei é essencial à validade do argumento, e as leis que fazem parte do *explanans* visivelmente constituem um fator relacionado à circularidade. Para suplementar esses desafios, relembremos o que Dretske salientou com relação às leis e explicações:

> Dizer que uma lei é uma verdade universal com poder explicativo é como dizer que uma cadeira é uma brisa no ar feita para sentar as pessoas. Você não consegue fazer uma bolsa de seda através da orelha de um porco, nem se fosse uma ótima orelha; e você não consegue fazer com que uma generalização explique suas instâncias, nem mesmo se fosse uma generalização puramente universal. O fato de todo F é G falhar em explicar por que qualquer F é G, isso se deve não porque seus esforços explicativos sejam muito fracos para chamarem nossa atenção, mas porque essa tentativa de explicação nunca é feita [...]. Subsumir uma instância sob uma generalização universal tem exatamente tanto poder explicativo quanto derivar Q de P&Q, nenhum. (1977, 26).

A reação de Dretske a essa citação foi concluir que as leis da natureza não são condicionais universalmente quantificadas; não são meras generalizações. Pelo contrário, pensou-se que as leis tivessem de ser algo de um tipo diferente: uma relação entre universais, generalizações, ou um axioma ou teorema de um sistema ideal, ou mesmo uma generalização metafisicamente necessária. Outras teorias devem ser levadas em consideração; mas talvez, quem sabe, as leis da natureza sejam apenas generalizações e não sejam explicativas em qualquer sentido significativo. Nesse caso teríamos uma explicação que identificaria qual tipo de entidade seria uma lei da natureza.

10. Física e as Ciências Especiais

Duas questões separadas, mas que estão relacionadas, têm recebido muita atenção recentemente na literatura ao redor das leis. Nenhuma delas têm muita relação com o que é uma lei, todavia, possuem uma relação com a natureza das generalizações que as/os cientistas tentam descobrir. Primeiro: alguma ciência tenta descobrir regularidades sem exceções no seu esforço de procurar leis? Segundo: mesmo que uma ciência, como a física fundamental, o faça, o que dizer quanto às outras?

10.1. A física tenta descobrir Regularidades sem Exceções?

A Filosofia traça uma distinção entre as generalizações **estritas** e as generalizações *ceteris-paribus*. O contraste deveria ser entre as generalizações universais do tipo proposto acima, como a de que todo corpo inercial não possua aceleração, e aparentemente as generalizações menos formais, tais como a de que, em circunstâncias iguais, fumar causa câncer. A ideia é que as primeiras seriam contraditas por uma simples contra-instância, isto é, um corpo inercial acelerando, entretanto, as segundas são consistentes com o fato de existir uma ou um fumante que nunca pega câncer. Embora na teoria essa distinção é suficientemente fácil de se entender, na prática é frequentemente mais difícil de distinguir a generalização estrita da *ceteris-paribus*. Isso ocorre porque muitas filósofas e muitos filósofos pensam que enunciados que não incluem uma cláusula *ceteris-paribus* explícita, implicitamente incluem tal cláusula.

Por muito tempo, filósofas e filósofos têm pensado que, se as/os cientistas já descobriram alguma regularidade sem exceções e que elas são leis, elas e eles o fizeram no nível da física fundamental. Uma quantidade significativa de filósofas e filósofos, no entanto, duvidam que haja regularidades sem exceções até mesmo no nível mais básico. Assim, por exemplo, Nancy Cartwright (1980, p. 75) argumentou que os aspectos descritivos e explicativos das leis conflitam, visto que “Fornecidas como descrições de fatos, elas são falsas; corrigidas para serem verdadeiras, elas perdem sua força explicativa fundamental.” Considere o princípio gravitacional de Newton, $\mathbf{F = Gmm'/r^2}$. Devidamente entendido, de acordo com Cartwright, ele diz

que para quaisquer dois corpos, a força entre eles é de $\mathbf{F = Gmm'/r^2}$. Mas se isso é o que a lei diz, então não é uma regularidade sem exceções. Isso porque a força entre dois corpos é influenciada por outras propriedades além apenas de suas massas e da distância entre eles, como por propriedades como a carga elétrica de dois corpos descrita na lei de Coulomb. A afirmação do princípio gravitacional pode ser corrigida para fazê-la verdadeira, mas isso, de acordo com Cartwright, pelo menos de acordo com a prática usual, poderia remover seu poder explicativo. Por exemplo, se o princípio é tomado como sustentando apenas que $\mathbf{F = Gmm'/r^2}$ se não há forças além da gravitacional atuando, então, embora fosse verdade, ele não se aplicaria exceto em circunstâncias idealizadas. Lange (1993) usa um exemplo diferente para apresentar um ponto similar. Considere a expressão padrão das leis de expansão térmica: "Sempre que a temperatura de uma barra de meta de comprimento $\mathbf{L_0}$ muda para **T**, o comprimento da barra muda por $\mathbf{L=kL_0T}$", onde *k* é uma constante, o coeficiente de expansão térmica do metal. Se essa expressão fosse usada para expressar a generalização estrita diretamente sugerida pela sua gramática, então tais enunciados seriam falsos, já que o comprimento da barra não muda do modo descrito em casos em que alguém está martelando as extremidades da barra. Parece que a lei requererá condições, mas tantas que o único jeito aparente de levar em consideração todas as condições requeridas seria por meio de uma cláusula *ceteris-paribus*. Então a preocupação é de que as afirmações sejam vazias. Por causa da dificuldade de afirmar condições plausíveis verdadeiras para sentenças *ceteris-paribus*, teme-se que "Em circunstâncias iguais, $\mathbf{L=kL_0T}$" pode apenas significar '$\mathbf{L=kL_0T}$, desde que $\mathbf{L=kL_0T}$'.

Mesmo aquelas e aqueles que concordam com o argumento de Cartwright e Lange, por vezes discordam sobre o que finalmente o argumento diz sobre as leis. Cartwright acredita que as leis verdadeiras não são regularidades sem exceções, mas, ao contrário, são afirmações que descrevem poderes causais. Se assim entendermos, as leis serão verdadeiras e explicativas. Lange acaba defendendo que há proposições propriamente adotadas como leis, embora, ao fazer isso, não é preciso também acreditar em alguma regularidade sem exceções. Giere (1999) pode ser proveitosamente interpretado como quem concorda com o argumento básico de Cartwright, mas insistindo que afirmações-lei não possuem condições implícitas ou cláusulas *ceteris-paribus* implícitas. Então, ele conclui que não existem leis.

Earman e Roberts asseguram que há regularidades sem exceções e com caráter de lei. Mais precisamente, eles argumentam que cientistas, ao fazerem física fundamental, tentam afirmar generalizações estritas são tais que seriam leis estritas se fossem verdadeiras:

> Nossa reivindicação é apenas que [...] teorias típicas para a física fundamental são tais que se elas fossem verdadeiras, haveria leis precisas livres de ressalvas. Por exemplo, a lei de campos gravitacionais de Einstein alega sem equívocos, qualificação, ressalva, cláusulas *ceteris-paribus* – que o tensor de curvatura de Ricci do espaço-tempo é proporcional ao total tensor de energia-momento pela energia-matéria; a versão relativista das leis de eletromagnetismo para o espaço-tempo plano livre de cargas de Maxwell alega – sem qualificações ou ressalva – que a curva do campo E é proporcional ao tempo parcial derivativo, etc. (1999, p. 446).

Sobre o exemplo gravitacional de Cartwright, eles pensam (473, fn. 14) que um entendimento plausível do princípio gravitacional é como descrever apenas uma força gravitacional entre dois corpos com massa (Cartwright argumenta que não há força componente e, assim, acha que tal interpretação seria falsa. Earman e Roberts discordam). Sobre o exemplo de Lange, eles pensam que leis deveriam ser entendidas como tendo a ressalva simples de que não há tensão externa na barra de metal (p. 461). Em todo caso, muito mais precisaria ser dito para estabelecer que toda generalização aparentemente estrita e explanatória que foi ou será estabelecida pelas físicas e pelos físicos se tornou ou se tornará falsa. A obra de Earman *et al*., de 2003, inclui artigos mais recentes de Cartwright e Lange, e também muitos outros artigos sobre leis *ceteris-paribus*.

10.2. Poderia haver alguma Lei de Ciência Especial?

Supondo que as físicas e os físicos tentem descobrir regularidades sem exceções, e até mesmo supondo que serão respectivamente bem sucedidas e bem sucedidos, às vezes, há a questão adicional de se é um objetivo de alguma ciência além da física fundamental, uma dita **ciência especial**, descobrir regularidades sem exceções e se essas/esses cientistas possuem alguma esperança de sucesso. Considere uma lei econômica de oferta e procura que diz que o preço cresce quando a procura aumenta e a oferta é mantida fixa. Note que, em alguns lugares, o preço

da gasolina tem, às vezes, se mantido o mesmo, apesar de um aumento na demanda e de uma oferta fixa, porque o preço da gasolina foi regulado pelo governo. Parece que a lei deve ser entendida como tendo uma cláusula *ceteris-paribus* para que ela seja verdadeira. Esse problema é bem geral. Como Jerry Fodor (1989, p. 78) já apontou, em virtude de ser declarada em um vocabulário de uma ciência especial, é muito provável que haja condições limitantes, especialmente condições físicas subjacentes, que enfraquecerão qualquer interessante generalização estrita de ciências especiais, condições que, elas mesmas, não poderiam ser descritas em um vocabulário de ciência especial. Donald Davidson incitou muito do recente interesse em leis de ciências especiais com seu *Mental Events* (1980, p. 207-225). Ele forneceu um argumento dirigido especialmente contra a possibilidade de leis estritas psicofísicas. Mais importante, ele sugeriu que a ausência de tais leis pode ser relevante para se eventos mentais causam eventos físicos. Isso levou uma série de artigos a lidarem com o problema de reconciliar a ausência de leis estritas de ciências especiais com a realidade de causação mental (LOEWER; LEPORE, 1987, 1989; FODOR, 1989; SCHIFFER, 1991; PIETROSKI; REY, 1995).

Progresso no problema das ressalvas depende da distinção entre três problemas básicos. Primeiro, há a questão do que é ser uma lei, que é, essencialmente, a busca por um complemento necessariamente verdadeiro para: "**P** é uma lei se e somente se [...]". Obviamente, para ser um complemento verdadeiro, deve-se assegurar para todo **P**, se **P** é uma generalização estrita ou uma *ceteris-paribus*. Segundo, há também a demanda de determinar as condições de verdade de sentenças de generalização usadas pelas/pelos cientistas. Terceiro, há uma questão, *a posteriori* e científica, de quais generalizações expressas por sentenças usadas pelas/pelos cientistas são verdadeiras. O terceiro desses problemas é onde a ação deve estar.

Sobre esse assunto, é impressionante quão pouca atenção é dada para os possíveis efeitos do contexto. Não poderia ser que, quando os economistas enunciam uma certa sentença de generalização estrita em uma "circunstância econômica" (ou seja, em uma apostila de economia ou em uma conferência de economia), considerações de sensitividade ao contexto afetando suas condições de verdade não tornariam esse enunciado verdadeiro? Poderia ser o caso, além do fato de que a mesma sentença enunciada em contextos diferentes (ou seja, em uma discussão entre físicas e físicos fundamentais ou, melhor ainda, em uma discussão filosófica de leis) pode resultar em um enunciado claramente falso. Essas mudanças de condições de verdade poderiam ser resultado de algo tão claro como uma troca contextual no domínio da quantificação ou talvez algo menos óbvio. O

que quer que seja, o ponto importante é que essa troca pode ser uma função de nada mais do que o significado linguístico da sentença e regras familiares de interpretação, a regra de acomodação.

Considere uma situação em que uma professora de engenharia profere: "Quando uma barra de metal é aquecida, a mudança no seu comprimento é proporcional à mudança na sua temperatura", e suponha que um estudante retruque: "Não quando alguém está martelando ambas as extremidades da barra". O estudante mostrou que o proferimento da professora era falso? Talvez não. Veja que o estudante terminou soando um pouco insolente. Em toda sua possibilidade, tal situação incomum, como alguém martelando em ambas as extremidades de uma barra quente, não entraria em jogo quando a professora disse o que disse. De fato, a razão pela qual o estudante terminou soando insolente é porque parece que ele deveria saber que seu exemplo era irrelevante. Note que a sentença da professora não precisa incluir alguma cláusula *ceteris-paribus* implícita para que seu proferimento seja verdadeiro. Como esse exemplo ilustra, em conversações simples, as velhas e claras sentenças de generalização estrita nem sempre são usadas para cobrir toda gama de casos. De fato, elas raramente são usadas desse jeito.

Se cientistas especiais proferem enunciados verdadeiros de sentenças de generalização, às vezes *ceteris-paribus*, então aparentemente nada as/os impede de enunciar sentenças verdadeiras nas ciências especiais quanto às leis. O problema aqui tem sido a verdade de generalizações de ciências especiais, não outro requerimento para algo ser uma lei.

11. Considerações Finais: o que vem em seguida?

Como tais questões progridirão? Como a filosofia poderia avançar para além das disputas atuais sobre as leis da natureza? Temos três questões especialmente interessantes e importantes; a primeira diz respeito ao problema de saber se as leis da natureza fazem parte das teorias científicas, questão esta levantada mais comumente em debates sobre causalidade do que sobre as leis. Roberts apresentou uma analogia para pensar que as leis não fazem parte das teorias científicas: é um postulado da geometria Euclidiana que dois pontos determinam uma linha, mas não é parte do conteúdo da geometria Euclidiana que esta proposição seja um postulado. A geometria Euclidiana não é uma teoria sobre postulados: trata-se de uma teoria acerca de pontos, linhas, e planos [...] (2008, p. 92). Pode ser que esse

seja um passo inicial em nossa compreensão sobre a ausência de termos nômicos em enunciados formais das teorias científicas. A segunda questão diz respeito ao problema de saber se há leis da natureza contingentes. A visão necessitarista tem sido cuidadosamente desenvolvida e regimentada, mas tais avanços receberam pouquíssima atenção por parte de certas filósofas e certos filósofos, como os humianos; é preciso mais estudo para explicar a fonte dos comprometimentos subjacentes que dividem esses campos. Por fim, é preciso ter mais atenção às linguagem utilizadas para reportar o que são as leis da natureza e a linguagem usada para expressar as leis propriamente ditas e se tais leis explicam. Não há dúvidas de que os embates recentes sobre as generalizações na física e nas ciências especiais giram exatamente em torno dessas questões, porém explorá-los também pode gerar bons frutos em questões centrais relacionadas à ontologia, realismo *versus* antirrealismo e sobreveniência.

Referências

ARMSTRONG, D. **A Theory of Universals**. Cambridge: Cambridge University Press, 1978.

ARMSTRONG, D. **What Is a Law of Nature?** Cambridge: Cambridge University Press, 1983.

ARMSTRONG, D. What Makes Induction Rational? **Dialogue**, v. 30, p. 503-511, 1991.

ARMSTRONG, D. The Identification Problem and the Inference Problem. **Philosophy and Phenomenological Research**, v. 53, p. 421-422, 1993.

BEEBEE, H. The Nongoverning Conception of Laws of Nature. **Philosophy and Phenomenological Research**, v. 61, p. 571-594, 2000.

BERENSTAIN, N.; LADYMAN, J. Ontic Structural Realism and Modality. *In:* LANDRY, E. RICKLES, D. (ed.). **Structural Realism:** Structure, Object, and Causality. Dordrecht: Springer, 2012.

BIGELOW, J.; ELLIS, B.; LIERSE, C. The World as One of a Kind: Natural Necessity and Laws of Nature. **British Journal for the Philosophy of Science**, v. 43, p. 371-388, 1992.

BIRD, A. The Dispositionalist Conception of Laws. **Foundations of Science**, v. 10, p. 353-370, 2005.

BLACKBURN, S. **Spreading the Word**. Oxford: Clarendon Press, 1984.

BLACKBURN, S. Morals and Modals. *In:* MACDONALD, G.; WRIGHT, C. (ed.). **Fact, Science and Morality**. Oxford: Basil Blackwell, 1986.

CARROLL, J. The Humean Tradition. **The Philosophical Review**, v. 99, p. 185-219, 1990.

CARROLL, J. **Laws of Nature**. Cambridge: Cambridge University Press, 1994.

CARROLL, J. (ed.). **Readings on Laws of Nature, Pittsburgh:** Pittsburgh University Press. Queen's University Press, 2004.

ELLIS, B.; LIERSE, C. Dispositional Essentialism. **Australasian Journal of Philosophy**, v. 72, p. 27-45, 1994.

FALES, E. **Causation and Universals**. London: Routledge, 1990.

FODOR, J. Making Mind Matter More. **Philosophical Topics**, v. 17, p. 59-79, 1989.

FOSTER, J. Induction, Explanation and Natural Necessity. **Proceedings of the Aristotelian Society**, v. 83, p. 87-101, 1983.

FOSTER, J. **The Divine Lawmaker**. Oxford: Clarendon Press, 2004.

FRIEND, T. Laws are Conditionals. **European Journal for the Philosophy of Science**, v. 6, p. 123-144, 2016.

GIERE, R. **Science Without Laws.** Chicago: University of Chicago Press, 1999.

GOODMAN, N. The Problem of Counterfactual Conditionals. **Journal of Philosophy**, v. 44, p. 113-128, 1947.

GOODMAN, N. **Fact, Fiction, and Forecast**. Cambridge: Harvard University Press, 1983.

HALL, N. **Humean Reductionism about Laws**. Oxford: John Wiley and Sons, 2015.

HEMPEL, C.; OPPENHEIM, P. Studies in the Logic of Explanation. **Philosophy of Science**, v. 15, p. 135-175, 1948.

HILDEBRAND, T. Can Primitive Laws Explain? **Philosophers Imprint**, v. 13, n. 5, 2013.

HILDEBRAND, T. Can Bare Dispositions Explain Categorical Regularities? **Philosophical Studies**, v. 167, n. 3, p. 569-584, 2014.

ISMAEL, J. How to be Humean. *In:* LEWIS, David, B. L. and SCHAFFER, J. (ed.). **A Companion to Oxford**: John Wiley and Sons, 2015.

JACKSON, F.; PARGETTER, R. Confirmation and the Nomological. **Canadian Journal of Philosophy**, v. 10, p. 415-428, 1980.

KRIPKE, S. **Naming and Necessity**. Cambridge: Harvard University Press, 1972.

LANGE, M. Natural Laws and the Problem of Provisos. **Erkenntnis**, v. 38, p. 233-248, 1993.

LANGE, M. **Natural Laws in Scientific Practice**. Oxford: Oxford University Press, 2000.

LANGE, M. A Note on Scientific Essentialism, Laws of Nature, and Counterfactual Conditionals. **Australasian Journal of Philosophy**, v. 82, p. 227-241, 2004,

LANGE, M. **Laws and Lawmakers**. New York: Oxford University Press, 2009.
LANGE, M. Grounding, Scientific Explanation, and Humean Laws. **Philosophical Studies**, v. 164, p. 255-261, 2013.
LANGE, M.; *et al*. Counterfactuals All the Way Down? Marc Lange: Laws and Lawmakers. **Metascience**, v. 20, p. 27-52, 2011.
LANGFORD, C. Review of An Interpretation of Causal Laws. **Journal of Symbolic Logic**, v. 6, p. 67-68, 1941.
LEWIS, D. **Counterfactuals**. Cambridge: Harvard University Press, 1973.
LEWIS, D. New Work for a Theory of Universals. **Australasian Journal of Philosophy**, v. 61, p. 343-377, 1983.
LEWIS, D. **Philosophical Papers**. v. II. New York: Oxford University Press, 1986.
LEWIS, D. Humean Supervenience Debugged. **Mind**, v. 103, p. 473-390, 1994,
LOEWER, B. Humean Supervenience. **Philosophical Topics**, v. 24, p. 101-126, 1996.
LOEWER, B. More on Making Mind Matter. **Philosophical Topics**, v. 17, p. 175-191, 1989.
LOEWER, B.; LEPORE, E. Mind Matters. **Journal of Philosophy**, v. 84, p. 630-642, 1987.
LYON, A. The Immutable Laws of Nature. **Proceedings of the Aristotelian Society**, v. 77, p. 107-126, 1976-1977.
MARSHALL, D. Humean Laws and Explanations. **Philosophical Studies**, v. 172, n. 12, p. 3145-3165, 2015.
MAUDLIN, T. **The Metaphysics Within Physics**. New York: Oxford University Press, 2007.
MILL, J. A. **System of Logic**. London: Longmans, Green and Co., 1947.
MILLER, E. Humean Scientific Explanation. **Philosophical Studies**, v. 172, n. 5, p. 1311-1332, 2015.
MUMFORD, S. **Laws in Nature**. London: Routledge, 2004.
PIETROSKI, P.; REY, G. When Other Things Aren't Equal: Saving Ceteris Paribus Laws from Vacuity. **British Journal for the Philosophy of Science**, v. 46, p. 81-110, 1995.
RAMSEY, F. **Foundations**. London: Routledge and Kegan Paul, 1978.
ROBERTS, J. Lewis, Carroll, and Seeing through the Looking Glass. **Australasian Journal ofPhilosophy**, v. 76, p. 426-438, 1998.
ROBERTS, J. **The Law-Governed Universe**. New York: Oxford University Press, 2008.
SCHAFFER, J. Causation and Laws of Nature: Reductionism. *In:* HAWTHORNE, J.; SIDER, T.; ZIMMERMAN, D. (ed.). **Contemporary Debates in Metaphysics**. Oxford: Basil Blackwell, 2008.
SCHIFFER, S. Ceteris Paribus Laws. **Mind**, v. 100, p. 1-17, 1991.

SCHNEIDER, S. What is the Significance of the Intuition that Laws of Nature Govern? **Australasian Journal of Philosophy**, v. 85, n.2, p. 307-324, 2007.

SHOEMAKER, S. Causality and Properties. *In:* VAN INWAGEN P. (ed.). **Time and Cause**. Dordrecht: D. Reidel Publishing Company, 1980.

SHOEMAKER, S. Causal and Metaphysical Necessity. **Pacific Philosophical Quarterly**, v. 79, p. 59-77, 1998.

SIDELLE, A. On the Metaphysical Contingency of Laws of Nature. *In:* GENDLER, T. S.; HAWTHORNE, J. (ed). **Conceivability and Possibility**. Oxford: Clarendon Press, 2002.

SOBER, E. Confirmation and Lawlikeness. **Philosophical Review**, v. 97, p. 93-98, 1988.

SWOYER, C. The Nature of Natural Laws. **Australasian Journal of Philosophy**, v. 60, p. 203-223, 1982.

TOOLEY, M. The Nature of Laws. **Canadian Journal of Philosophy**, v. 7, p. 667-698, 1977.

TOOLEY, M. **Causation**. Oxford: Clarendon Press, 1987.

TWEEDALE, M. **Armstrong on Determinable and Substantival Universals**. Dordrecht: D. Reidel Publishing Company, 1984.

VETTER, B. **Dispositional Essentialism and the Laws of Nature. Properties, Powers and Structures**. New York: Routledge, 2012.

VAN FRAASSEN, B. Armstrong on Laws and Probabilities. **Australasian Journal of Philosophy**, v. 65, p. 243-259, 1987.

VAN FRAASSEN, B. **Laws and Symmetry**. Oxford: Clarendon Press, 1989.

VAN FRAASSEN, B. Armstrong, Cartwright, and Earman on Laws and Symmetry. **Philosophy and Phenomenological Research**, v. 53, p. 431-444, 1993.

WARD, B. Humeanism without Humean supervenience: A projectivist account of laws and possibilities. **Philosophical Studies**, v. 107, p. 191-218, 2002.

WARD, B. Laws, Explanation, Governing, and Generation. **Australasian Journal of Philosophy**, v. 85, n. 4, p. 537-552.

WOODWARD, J. Realism about Laws. **Erkenntnis**, v. 36, p. 181-218, 2007.

Sobre os tradutores e revisores

Alex Bispo Bastos

Graduando em Filosofia pela Universidade Estadual de Campinas (Unicamp). Interessado em ética normativa, metaética e responsabilidade moral sob a perspectiva da filosofia analítica. Desenvolveu pesquisa de Iniciação Científica, financiada pela FAPESP (2020/09213-0), sobre o problema da existência de dilemas morais e a natureza da responsabilidade moral.

Ariston Sócrates G. S. Cruz

Graduando em Filosofia na UNICAMP. No momento, possui uma pesquisa em andamento sobre a semântica de demonstrativos pelo PIBIC (Processo 125600/2021-6).

Gabriel A. Eugênio

Graduando em Filosofia na UNICAMP e atualmente realiza pesquisa, com auxílio da FAPESP, sobre Ontologia dos Objetos Matemáticos (processo 2020/08214-3).

Gabriel Maruchi

Graduando em Filosofia pela UNICAMP. Atualmente realiza pesquisa em metafísica do livre-arbítrio pela FAPESP (Processo 2019/12207-5).

Karenina do Nascimento Rodrigues

Graduada e mestra em História pela Universidade Federal Rural do Rio de Janeiro-UFRRJ. Doutoranda em Teoria e História Literária pela Universidade Estadual de Campinas-UNICAMP.

Kherian Gracher

Doutor em Filosofia (2016 - 2020) pelo Programa de Pós-Graduação em Filosofia da Universidade Federal de Santa Catarina na área de Lógica e Epistemologia, orientado pelo Dr. Décio Krause e co-orientado pelo Dr. Newton Carneiro Affonso da Costa. Mestre em Filosofia (2014 - 2016) pela mesma instituição e também orientado pelo Dr. Décio Krause. Anteriormente, obteve o título de Bacharel em Filosofia pela Universidade Federal de Ouro Preto (2008 - 2013). Tem experiência em Filosofia Analítica, com ênfase em Metafísica, Filosofia da Lógica, Epistemologia, Filosofia da Ciência e História da Filosofia. Trabalha também com divulgação de Filosofia para o grande público, contribuindo para o portal Universo Racionalista como também sendo dono do canal de Youtube chamado "SciFilo".

Renato Semaniuc Valvassori

Graduando em Filosofia pela Universidade Estadual de Campinas (Unicamp). Interessado em metafísica contemporânea e filosofia antiga. Entre 2019-2020, desenvolveu pesquisa acerca do diálogo Hípias Maior, de Platão, sob financiamento do SAE da Unicamp. Mais recentemente, entre 2020-2021, realizou uma pesquisa financiada pela FAPESP sobre ontologia de mundos possíveis

Rhamon de Oliveira Nunes

Bacharel em filosofia pela Universidade Federal Fluminense (UFF), mestre e doutor em filosofia pelo Programa de Pós-Graduação em Lógica e Metafísica (PPGLM) da Universidade Federal do Rio de Janeiro (UFRJ). Sua tese de doutorado recebeu o Prêmio CAPES 2021. Foi professor substituto na mesma instituição nos anos de 2018 e 2019. Atualmente é professor temporário no departamento de filosofia da Universidade do Estado do Rio Grande do Norte (UERN), ocupando a cadeira de Filosofia Antiga e Medieval. Sua pesquisa primária tem como foco a mereologia, a saber, o estudo lógico-ontológico dos conceitos de parte e todo. Interessa-se também por outros temas em metafísica contemporânea, tais quais: filosofia do espaço, filosofia do tempo, identidade diacrônica e fundação metafísica. Sua pesquisa secundária aborda os fundamentos da leitura, escrita de textos de filosofia bem como a própria natureza da metodologia filosófica.

Ulisses Franceschi Eliano

Graduando em Filosofia pela UNICAMP, estuda temas como abdução, explicação científica, paraconsistência e lógica.

Vinicius Bianchi

Possui licenciatura (2017) e mestrado (2019) em filosofia, ambos pela Universidade Federal de Pelotas, RS. Realizou pesquisa de iniciação científica acerca da responsabilidade moral e agência sob orientação do Prof.Dr. João Hobuss que também atuou como orientador durante o mestrado. Defendeu sua dissertação com o título " O problema da vontade livre: Uma abordagem contemporânea" e atualmente é doutorando em filosofia pela Universidade Federal de Santa Catarina, SC, onde pesquisa as condições de responsabilidade moral vinculadas à epistemologia sob orientação do Prof.Dr. Alexandre Meyer Luz.

Editora
UFPel

DISSERTATIO
FILOSOFIA