Editora
UFPel

Série Investigação Filosófica

Textos selecionados de

Filosofia da Ciência II

Luana Poliseli
(Organizadora)

DISSERTATIO
FILOSOFIA

TEXTOS SELECIONADOS DE FILOSOFIA DA CIÊNCIA II

Série Investigação Filosófica

TEXTOS SELECIONADOS DE FILOSOFIA DA CIÊNCIA II

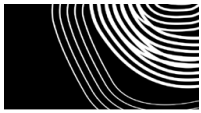
Luana Poliseli
(Organizadora)

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)

ORGANIZADORA DO VOLUME

Luana Poliseli (KLI))

TRADUTORES E REVISORES

Ana Margarete Barbosa de Freitas (UFBA)
Felipe Rocha Lima Santos (UFBA)
Jessica Suellen Oliveira Costa (UFBA)
Lúcia Carvalho Neco (LTU)
Mariana Vitti Rodrigues (USP)
Nathalia Cristina Alves Pataleão Strongen (UNICAMP)
Pedro Bravo de Souza (USP)
Renata Arruda (UFG)
Tiago Luís Teixeira de Oliveira (UERJ)

CRÉDITO DA IMAGEM DE CAPA

Joseph Wright of Derby, “An experimente on a Bird in an Air Pump” de 1768.
Fonte: 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 filosofia da ciência II.
[recurso eletrônico] Organizadora: Luana Poliseli – Pelotas: NEPFIL Online, 2021.

422p. - (Série Investigação Filosófica).

Modo de acesso: Internet
<wp.ufpel.edu.br/nepfil>
ISBN: 978-65-86440-36-2

1. Filosofia. 2. Ciência I. Poliseli, Luana.

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 **16**
Introdução **18**

(I) Explicação Científica **21**
1. Antecedentes e Introdução 22
2. O Modelo ND 25
2.1 A Ideia Básica 25
2.2 O Papel das Leis no Modelo ND 26
2.3 Explicação Indutiva Estatística 29
2.4 Motivação para o Modelo ND: Expectabilidade Nômica e uma Formulação de Regularidade da Causalidade 30
2.5 Entendimento Explicativo e Expectabilidade Nômica: Contraexemplos à Suficiência 34
2.6 A Estratégia de Estrutura Oculta 36
3. O Modelo RE 40
3.1 A Ideia Básica 40
3.2 O Modelo RE e Eventos de Baixa Probabilidade 42
3.3 O que Teorias Estatísticas Explicam? 45
3.4 Relações de Causalidade e Relevância Estatística 49
4. O modelo Mecânico-Causal 51
4.1 A Ideia Básica 51
4.2 O Modelo MC e Relevância Explicativa 53
4.3 O Modelo MC e Sistemas Complexos 56
4.4 Desenvolvimentos Mais Recentes 60
5 Uma Abordagem Unificacionista de Explicação 62
5.1 A Ideia Básica 62
5.2 Ilustrações do Modelo Unificacionista 64

5.3 As ilustrações Criticadas 65
5.4 A Heterogeneidade da Unificação 67
5.5 A Concepção O-Vencedor-Leva-Tudo de Unificação Explicativa 69
5.6 A Epistemologia da Unificação 71
6 Teorias Pragmáticas da Explicação 74
6.1 Introdução 74
6.2 Empirismo Construtivo e Teoria Pragmática da Explicação 80
6.3 Explicando Como um Ato Ilocucionário 86
6.4 Reflexões Finais sobre Teorias Pragmáticas 89
7 Conclusões, Questões em Aberto e Direções Futuras 92
7.1 O Papel da Causalidade 92
7.2 Explicação e Outros Objetivos Epistêmicos 95
7.3 Um Único Modelo de Explicação? 97
Referências 99

(II) Descoberta Científica 103
1. Introdução 104
2 A Investigação Científica Como Descoberta 105
3 Elementos da Descoberta 107
4 Lógicas da Descoberta 110
5 Distinção Entre Contexto de Descoberta e Contexto de Justificação 111
6 A Lógica da Descoberta Depois da Distinção de Contexto 116
6.1 Descoberta Como Abdução 116
6.2 Programação Heurística 119
7 Anomalias e a Estrutura da Descoberta 121
8 Metodologias de Descoberta 123
8.1 Descobertabilidade 124
8.2 Avaliação Preliminar 125
9. Criatividade, Analogia e Modelos Mentais 127
9.1 Condições Psicológicas e Sociais da Criatividade 128
9.2 Analogias 129
9.3 Modelos Mentais 130
Referências 132

(III) Abdução **138**
1 Abdução: Uma Noção Geral 139
1.1 Dedução, Indução, Abdução 140
1.2 A Ubiquidade da Abdução 142
2 Explicando a Abdução 146
3 O Status da Abdução 152
3.1 Críticas 152
3.2 Defesas 154
4 Abdução Versus Teoria Bayesiana da Confirmação 157
Suplemento: Abdução Peirciana 162
Referências 165

(IV) Predição versus Acomodação **172**
1 Introdução Histórica 173
2 Hipóteses *ad hoc* 176
3 Caracterizações Iniciais de Novidade 178
4 Uma Taxonomia Preditivista 181
5 A Tese do Suporte Nulo 181
6 Teorias Contemporâneas do Preditivismo 183
6.1 Métodos de Descoberta Confiáveis 183
6.2 A Explicação Forjada 185
6.3 Conjunções Arbitrárias e Não Arbitrárias 186
6.4 Testes Severos 188
6.5 Confirmação Condicional e Incondicional 191
6.6 A Analogia do Arqueiro 193
6.7 A Abordagem de Akaike 195
6.8 A Novidade de Endosso e a Confirmação de Crenças de Fundo 197
7 Antipreditivismo 200
8 O Debate Realista/Antirrealista 202
Referências 206

(V) Representação Científica **217**
1 Problemas Relacionados a Representação Científica 218
2 Griceanismo Geral e Decreto Estipulativo 226

3 A Concepção de Similaridade 229
3.1 Similaridade e Problema RE 230
3.2 Precisão, Estilo e Ontologia 234
4 A Concepção Estruturalista 236
4.1 Estruturas e o Problema da Ontologia 236
4.2 Estruturalismo e o Problema RE 238
4.3 Demarcação, Precisão, Estilo e Estrutura Alvo-Final 241
5 A Concepção Inferencial 244
5.1 A Explicação DDI 244
5.2 Inferencialismo Deflacionário 245
5.3 Inflando o Inferencialismo: Interpretação 246
6 A Visão Ficcional dos Modelos 248
7 Representação-Como 251
7.1 Da Arte a Ciência 251
7.2 A Explicação DEKI 253
Referências 255

(VI) Modelos na Ciência 265
1 Semântica: Modelos e Representação 266
2 Ontologia: O que são Modelos? 275
2.1 Objetos Físicos 275
2.2 Objetos Fictícios e Objetos Abstratos 276
2.3 Estrutura de Teorias dos Conjuntos 278
2.4 Descrições e Equações 279
3 Epistemologia: as Funções Cognitivas dos Modelos 281
3.1 Aprendendo Sobre Modelos 281
3.2 Aprendendo Sobre Sistemas-Alvo 283
3.3 Explicando com Modelos 283
3.4 Compreensão com Modelos 285
3.5 Outras Funções Cognitivas 287
4 Modelos e Teoria 288
4.1 Modelos como Subsidiários de Teorias 288
4.2 Modelos como Independentes de Teorias 289
5 Modelos e Outros Debates na Filosofia da Ciência 294
5.1 Modelos, Realismo e Leis da Natureza 294

5.2 Modelos e Reducionismo 297
Referências 288

(VII) Mecanismos na Ciência 317
1 A Ascensão do Novo Mecanismo 318
2 O Conceito de Mecanismo 319
2.1 Fenômeno 322
2.1.1 Produzir, Subjazer e Manter 323
2.1.2 Regularidade 324
2.2 Partes 325
2.3 Causas 326
2.3.1 Abordagens de Quantidade Conservada 326
2.3.2 Abordagens Mecanísticas 327
2.3.3 Abordagens Baseadas em Atividades 328
2.3.4 Abordagens Contrafactuais 328
2.4 Organização 329
2.4.1 Organização e Agregatividade 329
2.4.2 Variedades de Organização 330
2.4.3 Modularidade 331
2.4.4 Junção 332
2.4.5 Níveis 332
2.4.6 Mecanismos Estáveis e Efêmeros 333
2.5 O Que Mecanismos Não São e o Que Não São Mecanismos 333
2.5.1 O Que Mecanismos Não São 333
2.5.2 O Que Não São Mecanismos 336
2.6 Trabalho Filosófico a Ser Feito 338
3 Explicação: das Análises Formais às Estruturas Materiasis 339
3.1 Explicações Etiológicas e Constitutivas 340
3.2 Relevância Constitutiva 342
3.3 Mecanismos e Modelos 343
3.3.1 Caracterizando a Completude 344
3.3.2 Abstração e Idealização 345
3.4 Trabalho Filosófico a Ser Feito 347
4 Metafisica dos Mecanismos 348
4.1 Mecanismos de Instância/Espécime, Mecanismos de Tipo e Leis 348

4.2 Mecanismos, Níveis e Emergência 349
4.3 Mecanismos e Realização 351
4.4 Mecanismos e Tipos Naturais 352
4.5 Mecanismos e Funções 353
4.6 Trabalho Filosófico a Ser Feito 354
5 Relações Entre Disciplinas Científicas: da Teoria da Redução à Integração do Mecanismo 354
5.1 Teoria da Redução 354
5.2 Integração do Mecanismo 355
5.2.1 Pluralismo Integrativo 356
5.3 Trabalho Filosófico a Ser Feito 357
6 Descoberta: dos Momentos Eureka às Estratégias de Descoberta 357
6.1 Descoberta via Momentos Eureka 357
6.2 Descoberta via Estratégias 358
6.3 Evidência Mecanistica em Descobertas Médicas 361
6.4 Trabalho Filosófico a Ser Feito 362
7 Conclusão 362
Referências 363

(VIII) Simulações Computacionais na Ciência 379
1 O Que é Simulação Computacional? 380
1.1 Uma definição estrita 380
1.2 Uma definição geral 381
1.3 Um ponto de vista alternativo 382
2 Tipos de Simulações Computacionais 384
2.1 Simulações Baseadas em Equações 385
2.2 Simulações Baseadas em Agentes 385
2.3 Simulações de Múltiplas Escalas 386
2.4 Simulações de Monte Carlo 388
3 Propósitos da Simulação 389
4 A Epistemologia das Simulações Computacionais 390
4.1 Novos Recursos da ESC 391
4.2 ESC e a Epistemologia do Experimento 393
4.3 Verificação e Validação 395
4.4 ESC e Direito Epistêmico 398

4.5 Abordagens Pragmáticas da ESC 399
5 Simulações e Experiência 400
6 Simulações Computacionais e a Estrutura das Teorias Cientificas 406
7 Emergência 408
8 Ficções 408
Referências 411

Sobre Tradutores e Revisores **416**

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

O *volume I de textos selecionados em Filosofia da Ciência*, organizado pelo Dr. Tiago Luis Teixeira de Oliveira, presenteia-nos com uma bela introdução acerca dos aspectos epistemológicos, metafísicos e éticos desta disciplina que tem por objetivo estudar a ciência. Através de uma meticulosa escolha entre os mais diversos verbetes da *Stanford Encyclopedia of Philosophy* (como por exemplo, Ciência e Pseudociência, Método Científico, Realismo Científico, Objetividade Científica, Dimensão Social do Conhecimento Científico, e Filosofia da Tecnologia), temos, em suas palavras, no volume I, uma "apresentação da natureza do conhecimento científico e de seus valores".

Dando continuidade a este intento, o presente livro, *volume II de textos selecionados em Filosofia da Ciência*, pretende, por sua vez, trazer aos leitores (estudantes, professores, pesquisadores e curiosos) verbetes que discutem temas relacionados ao desenvolvimento da atividade científica – ou em outras palavras, como a prática científica acontece. Pode parecer trivial, mas perguntas que, em um primeiro momento, parecem simples como *O que são explicações científicas? Como os cientistas explicam os fenômenos? Para que servem os modelos? O que são mecanismos? O que são simulações computacionais e pra que servem? É tudo, em ciência, preditivo?* e *O que é uma descoberta científica*? estão carregadas de imenso valor filosófico que, geralmente, acompanham calorosos (e por vezes inconclusivos) debates entre os estudiosos desta área, e que são responsáveis por fornecerem os arcabouços teóricos pelo qual a ciência é construída.

Pensando em convidar os leitores a uma imersão nestes debates filosóficos e epistemológicos que subsidiam e fornecem os pilares para a ciência, este livro apresenta os seguites verbetes: (I) Explicação Científica, (II) Descoberta Científica, (III) Abdução, (IV) Predição versus Acomodação, (V) Representação Científica, (VI) Modelos na Ciência, (VII) Mecanismos na Ciência, e (VIII) Simulações Computacionais na Ciência.

Começamos este livro com o capítulo *Explicação Científica*. A ideia é, logo de início, apresentar aos leitores os diversos tipos de explicações que são encontrados na ciência, como eles se diferenciam dos contextos não-científicos; trazer algumas objeções à esses modelos explicativos; e também, discutir como as teorias se relacionam com as explicações. Logo na sequência, trazemos o capítulo *Descoberta Científica*, cujo foco é apresentar aos leitores a diferença entre o que é o resultado da ciência do que é o procedimento da investigação científica. A distinção entre o *contexto de descoberta* (os processos de pensamento *de fato*) e o *contexto da justificativa* (a defesa *de jure* da correção desses pensamentos) dominou e moldou a filosofia da ciência do século XX, e para aqueles leitores interessados em compreender como a ciência opera, faz-se crucial entender a distinção entre a geração de uma nova ideia (ou hipótese) da sua defesa (teste, verificação). Intimamente relacionado ao verbete anterior, passamos agora ao capítulo *Abdução,* onde o cerne está no raciocínio explanatório durante a *geração* de hipóteses e a *justificação* de hipóteses (sendo este último sentido, comumente referido como *inferência a melhor explicação*, e também, o mais frequentemente usado). Este verbete mostra que a abdução é um tipo de inferência empregada tanto no raciocínio cotidiano quanto no científico, e contrasta a abdução com os outros tipos de inferências explicativas – mostrando seus usos relevantes e caracterizações normativas. O capítulo *Predição versus Acomodação* cumpre o papel de mostrar o debate sobre as avaliações epistêmicas das teorias e por consequência seus respectivos sucessos. Seriam as teorias bem-sucedidas ao gerarem predições especificas sobre os fenômenos (predição) ou somente por passarem pela verificação (acomodação)? No capítulo *Representação Científica* o foco é mostrar que as ideias científicas podem ser representadas por diversas formas (sejam elas teorias, modelos, gráficos, imagens, equações, asserções linguísticas, etc.) mas que essas representações enfretam problemas diversos que precisam ser adequadamente direcionadas pelos cientistas, como por exemplo, o que transforma algo em uma representação científica de alguma outra coisa?; Qual o papel desta representação?; e, como uma representação científica se relaciona com uma explicação cientítfica? No quesito representações, um dos grandes fortes da ciência é o desenvolvimento de modelos, o capítulo *Modelos na Ciência* nos direciona às diversas perguntas que nutrem esta temática, são questões semânticas (como os modelos representam e se é que representam), ontológicas (que tipo de coisas são modelos), e epistemológicas (como aprendemos e explicamos os modelos). Uma forma bastante frequente de explicar e representar fenômenos na ciência é através do uso de

mecanismos, o capítulo *Mecanismos na Ciência* se dedica a mostrar como esta forma de pensamento reorientou os trabalhos filosóficos. Neste verbete, uma introdução e resumo das suas características distintivas são apresentadas, sobretudo, sua relação com as explicações, com os debates metafísicos, suas interações com as diversas disciplinas científicas bem como com o processo de descoberta científica. Por último, finalizamos o livro com o verbete *Simulações Computacionais na Ciência*. Este capítulo aborda questões sobre a estrutura epistemológica da simulação computacional, quais suas relações com experimentos e estruturas científicas, entre outros.

Os verbetes supracitados foram traduzidos e revisados pela Dra. Ana Margarete Barbosa Ferreira, Dr. Felipe Rocha Lima Santos, Me. Jessica Suellen Oliveira Costa, Me. Lúcia Carvalho Neco, Dra. Mariana Vitti Rodrigues, Dra. Nathalia Cristina Alves Pantaleão, Me. Pedro Bravo de Souza, Dra. Renata Maria Santos Arruda, e Dr. Tiago Luis Teixeira de Oliveira – sem os quais este livro não poderia vir a ser. Portanto, a prazerosa leitura deste volume é em sua íntegra devido aos primorosos esforços desses tradutores e revisores, bem como aos organizadores gerais da *Série Investigações Filosóficas* Dr. Rodrigo Reis Lastra Cid e Dr. Juliano do Carmo, além do Dr. Edward Zaltar da *Stanford Encyclopedia of Philosophy* que viabilizaram esta iniciativa.

A *todes*, uma salva de palmas.

Dra. Luana Poliseli
Organizadora do Livro

(I) Explicação Científica*

Autor: James Woodward
Tradução: Felipe Rocha L. Santos
Revisão: Pedro Bravo

Questões relativas à explicação científica têm sido um foco de atenção filosófica desde os tempos pré-socráticos até o período moderno. No entanto, a discussão recente realmente começa com o desenvolvimento do modelo Nomológico-Dedutivo (ND). Este modelo teve muitos defensores (POPPER, 1935, 1959; BRAITHWAITE, 1953; GARDINER, 1959; NAGEL, 1961), mas, sem dúvida, a declaração mais detalhada e influente se deve a Carl Hempel (1942, 1965; HEMPEL & OPPENHEIM, 1948). Esses artigos, e a reação a eles, estruturaram a discussão subsequente sobre a explicação científica em um grau extraordinário. Após algumas observações gerais por meio de um plano de fundo histórico e orientação (Seção 1), esta entrada descreve o modelo ND e suas extensões e, em seguida, volta para

---

* WOODWARD, J. Scientific Explanation. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2019. Disponível em: http://plato.stanford.edu/archives/win2019/entries/scientific-explanation/. Acesso em: 11 ago. 2021.

The following IE the translation of the entry on Scientific Explanation by James Woodward, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at http://plato.stanford.edu/archives/win2019/entries/scientific-explanation/. 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 IE located at http://plato.stanford.edu/entries/scientific-explanation/ 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.

algumas objeções bem conhecidas (Seção 2). Posteriormente, descreve uma variedade de tentativas subsequentes de desenvolver modelos alternativos de explicação, incluindo os modelos de Relevância Estatística de Wesley Salmon (Seção 3) e Mecânica Causal (Seção 4) e os modelos Unificacionistas de Michael Friedman e Philip Kitcher (Seção 5). A seção 6 fornece um resumo e discute orientações para trabalhos futuros.

1. Antecedentes e Introdução

Como ficará aparente, *explicação científica* é um tópico que levanta uma série de questões inter-relacionadas. Alguma orientação em segundo plano será útil antes de se voltar para os detalhes dos modelos concorrentes. Um pressuposto da discussão mais recente tem sido o de que a ciência às vezes fornece explicações (em vez de algo que fica aquém da explicação – por exemplo, "mera descrição") e que a tarefa de uma "teoria" ou "modelo" de explicação científica é caracterizar a estrutura de tais explicações. Supõe-se, portanto, que exista (em algum nível de descrição adequadamente abstrata e geral) um único tipo ou forma de explicação que seja "científica". De fato, a noção de "explicação científica" sugere pelo menos dois contrastes - primeiro, um contraste entre as "explicações" que são características da "ciência" e as explicações que não são, e, segundo, um contraste entre "explicação" e alguma outra coisa. No entanto, com relação ao primeiro contraste, a tendência em grande parte da literatura filosófica recente tem sido assumir que existe uma continuidade substancial entre os tipos de explicações encontradas na ciência e pelo menos algumas formas de explicação encontradas em contextos não-científicos mais ordinários, em que estas incorporam de maneira mais ou menos incipientes propriedades que estão presentes naqueles de forma mais detalhada, precisa, rigorosa, etc. Supõe-se ainda que é tarefa de uma teoria da explicação capturar o que é comum às formas de explicações tanto científicas como mais ordinárias. Essas suposições ajudam a explicar (o que de outra forma pode parecer curioso ao leitor) porque, como este texto ilustrará, discussões sobre explicação científica frequentemente andam de um lado para o outro entre exemplos extraídos da ciência genuína (por exemplo, explicações das trajetórias dos planetas que apelam à mecânica newtoniana) e exemplos mais caseiros que envolvem a queda de tinteiros.

No que diz respeito ao segundo contraste, a maioria dos modelos de explicação assumem que seja possível que um conjunto de alegações seja verdadeiro,

preciso, apoiado em evidências e assim por diante e, no entanto, não seja explanatório (pelo menos qualquer coisa que o típico buscador-de-explicação provavelmente quer explicado). Por exemplo, todos os projetos de explicações científicas descritos abaixo concordariam que um relato da aparição de uma espécie específica de ave do tipo encontrado em um guia de aves não é, por mais preciso que seja, uma explicação de algo de interesse para os biólogos (por exemplo, o desenvolvimento, características ou comportamento dessa espécie). Ao invés disso, esse relato é "meramente descritivo". No entanto, diferentes modelos de explicação fornecem projetos diferentes sobre em que consiste o contraste entre o explicativo e o meramente descritivo.

Um ponto relacionado é que, embora a maioria dos teóricos da explicação científica tenha proposto modelos que visam cobrir pelo menos alguns casos de explicação que não consideraríamos como parte da ciência, eles ainda assim assumiram alguma restrição implícita aos tipos de explicação que eles buscaram reconstruir. Observou-se frequentemente que a palavra “explicação” é usada em uma ampla variedade de maneiras no português comum - falamos de explicar o significado de uma palavra, explicar o contexto das teorias filosóficas da explicação, explicar como assar uma torta, explicar por que alguém tomou uma certa decisão (onde isso é o mesmo que oferecer uma justificativa) e assim por diante. Embora os vários modelos discutidos abaixo às vezes tenham sido criticados por não conseguirem capturar todas essas formas de "explicação" (*vide* SCRIVEN, 1959), é claro que eles nunca tiveram esta intenção. Em vez disso, o *explicandum* pretendido é, grosso modo, explicações de *porque* as coisas acontecem, onde as “coisas” em questão podem ser eventos particulares ou algo mais geral - por exemplo, regularidades ou padrões repetíveis na natureza. Paradigmas desse tipo de explicação incluem a explicação para o avanço no periélio de mercúrio fornecida pela Relatividade Geral, a explicação da extinção dos dinossauros em termos do impacto de um grande asteroide no final do período Cretáceo, a explicação fornecida pela polícia de porque ocorreu um acidente de trânsito (o motorista estava bebendo e havia gelo na estrada), e a explicação padrão fornecida nos livros de economia sobre porque os monopólios, em comparação com empresas em mercados perfeitamente competitivos, aumentam os preços e reduzem a produção.

Finalmente, algumas palavras sobre o contexto epistemológico/metodológico mais amplo dos modelos descritos abaixo. Muitos filósofos pensam em conceitos como "explicação", "lei", "causa" e "apoio a contrafactuais" como parte de uma família inter-relacionada ou círculo de conceitos de caráter "modal". Por razões

"empiristas" familiares, Hempel e muitos outros defensores iniciais do modelo ND consideraram esses conceitos como pouco compreendidos, pelo menos antes da análise. Assumia-se que seria “circular” explicar um conceito dessa família em termos de outros da mesma família e que eles deveriam ser explicados em termos de outros conceitos de fora da família modal - conceitos que obviamente satisfaziam (o que foram considerados) padrões empiristas de inteligibilidade e testabilidade. Por exemplo, na versão do de Hempel do modelo ND, a noção de “lei” desempenha um papel fundamental na explicação do conceito de “explicação”, e sua suposição é de que leis são apenas regularidades que atendem a certas condições adicionais que também são aceitáveis para empiristas. Como veremos, esses padrões empiristas (e a falta de vontade de empregar conceitos modais como primitivos) continuaram a desempenhar um papel central nos modelos de explicação desenvolvidos subsequentemente ao modelo ND.

Há muitas questões históricas interessantes sobre o modelo ND que permanecem praticamente inexploradas. Por que, quando surgiu, 'explicação científica' surgiu como um tópico importante para a discussão filosófica? Por que os filósofos da ciência empiristas lógicos, que defendiam o modelo ND, estavam tão dispostos a aceitar a ideia de que a ciência fornece "explicações", dada a tendência de muitos escritores anteriores da tradição positivista de pensar em "explicação" como uma questão bastante subjetiva ou "metafísica" e contrastá-la desfavoravelmente com "descrição", a qual eles consideravam um objetivo mais legítimo para a ciência empírica? E por que a discussão, ao menos inicialmente, foi organizada em torno de "explicação" em vez de "causação", já que (como veremos) é frequentemente a última noção que parece ser de interesse central nos debates subsequentes e uma vez que a noção anterior parece (para muitas sensibilidades contemporâneas) um tanto vaga e mal definida? Pelo menos parte da resposta a esta última pergunta parece ser que (novamente, como será explicado em mais detalhes abaixo) Hempel e outros defensores do modelo ND herdaram padrões empiristas ou escrúpulos humeanos sobre a noção de causalidade. Eles assumiram que as noções causais são apenas (cientifica ou metafisicamente) aceitáveis na medida em que é possível parafrasear ou reformulá-las de maneira a satisfazer os critérios empiristas de significado e legitimidade. Uma maneira óbvia de fazer isso era considerar afirmações causais equivalentes a afirmações sobre a obtenção de “regularidades” (isto é, padrões de associação uniforme na natureza). É apenas essa ideia que é capturada pelo modelo ND (veja abaixo). Parte do apelo inicial do tópico "explicação científica" era, portanto, que ele funcionava como um

substituto mais respeitável (ou ponto de entrada) para o tópico problemático da causalidade[1] .Outra motivação foi o interesse de Hempel e de outros defensores iniciais do modelo de ND em formas de explicação como "explicação funcional" (pensada para ser empregada em ciências especiais como biologia e antropologia) que não eram obviamente causais. Isso também tornou natural enquadrar a discussão em torno de uma ampla categoria de explicação, em vez de noções mais restritas de "causalidade" (HEMPEL, 1965b).

Salmon (1989) traz um excelente levantamento crítico de todos os modelos de explicação científica discutidos neste texto. Pitt (1988) e Ruben (1993) são antologias que contêm vários artigos influentes.

2. O modelo ND

2.1 A ideia básica

De acordo com o Modelo Nomológico-Dedutivo, uma explicação científica consiste em dois "constituintes" principais: um *explanandum*, uma frase "descrevendo o fenômeno a ser explicado" e um *explanans*, "a classe daquelas frases que são aduzidas para explicar o fenômeno" (HEMPEL; OPPENHEIM, 1948, reimpresso em HEMPEL, 1965, p. 247). Para que o *explanans* possam explicar com sucesso o *explanandum*, várias condições devem ser atendidas. Primeiro, "o *explanandum* deve ser uma consequência lógica do *explanans*" e "as sentenças que constituem o *explanans* devem ser verdadeiras" (HEMPEL, 1965, p. 248). Ou seja, a explicação deve assumir a forma de um argumento dedutivo sólido, no qual o *explanandum* segue como conclusão a partir das premissas do *explanans*. Este é o componente "dedutivo" do modelo. Segundo, o *explanans* deve conter pelo menos uma "lei da natureza" e isso deve ser uma premissa essencial na derivação no sentido de que a derivação do *explanandum* não seria válido se essa premissa fosse removida. Esse é o componente "nomológico" do modelo - "nomológico" é um termo filosófico técnico que, suprimindo algumas sutilezas, significa (aproximadamente) "legal". Em

---

[1] *Vide* Cartwright (2004) para um diagnóstico semelhante e para outra pesquisa de alguns dos problemas descritos aqui.

sua formulação mais geral, o modelo ND deve aplicar-se tanto à explicação de "regularidades gerais" ou "leis" como (para usar os exemplos de Hempel e Oppenheim) porque a luz está em conformidade com a lei da refração e também com a explicação de eventos particulares, concebidos como ocorrendo em um determinado momento e local, como a aparência dobrada dos remos parcialmente submersos de um barco em uma ocasião específica de visualização. Como uma ilustração adicional de uma explicação ND de um evento específico, considere uma derivação da posição de Marte em algum momento futuro a partir das leis do movimento de Newton, a lei newtoniana do inverso do quadrado que governa a gravidade e informações sobre a massa do Sol, a massa de Marte e a posição atual e velocidade de cada um. Nesta derivação, as várias leis newtonianas figuram como premissas essenciais e são usadas, em conjunto com informações apropriadas sobre as condições iniciais (as massas de Marte e do sol e assim por diante), para derivar o *explanandum* (a futura posição de Marte) via um argumento dedutivamente válido. Os critérios ND são, então, satisfeitos.

2.2 O papel das leis no modelo ND

A noção de um argumento dedutivo sólido é (sem dúvida) relativamente clara (ou pelo menos algo que pode ser considerado como previamente compreendido do ponto de vista da caracterização da explicação científica). Mas e o outro componente principal do modelo de ND – o de uma lei da natureza? A intuição básica que guia o modelo de ND é mais ou menos assim: dentro da classe de verdadeiras generalizações, podemos distinguir entre aquelas que são apenas "acidentalmente verdadeiras" e aquelas que são "leis". Para usar os exemplos de Hempel, a generalização

(2.2.1) Todos os membros do Greensbury School Board de 1964 são carecas

é, se for verdade, apenas acidentalmente. Em contraste,

(2.2.2) Todos os gases se expandem quando aquecidos sob pressão constante

é uma lei. Assim, de acordo com o modelo ND, a última generalização pode ser usada, em conjunto com informações de que uma amostra específica de gás foi aquecida sob pressão constante, para explicar por que ela se expandiu. Por outro lado, a primeira generalização (2.2.1), em conjunto com a informação de que uma pessoa em particular **n** é membro do conselho escolar de Greensbury em 1964, não pode ser usada para explicar por que **n** é careca.

Embora esse exemplo possa parecer bastante claro, o que exatamente distingue verdadeiras generalizações acidentais de leis? Este tem sido objeto de muitas discussões filosóficas, a maioria das quais deve estar além do escopo deste texto[2] . Por razões explicadas na Seção 1, Hempel assume que um relato adequado deve explicar a noção de lei em termos de noções que estão fora da família modal[3] . Em Hempel (1965), ele considera várias propostas familiares com esse caráter[4] e conclui que todas são insuficientes, observando que o problema de caracterizar a noção de lei se mostrou "altamente recalcitrante" (1965, p.338). Parece justo dizer, no entanto, que sua suposição subjacente é que, no fundo, as leis são apenas generalizações sem exceção que descrevem regularidades que atendem a certas condições distintas adicionais que ele, no momento, não é capaz de formular.

Nas décadas subsequentes, houve vários outros critérios propostos para o estatuto de lei. Embora cada proposta tenha seus adeptos, nenhuma ganhou aceitação geral[5] . Que implicações isso tem para o modelo ND? Uma avaliação possível é que tudo o que o modelo de ND realmente requer é que haja concordância

---

[2] Além de Hempel (1965) e Salmon (1989), ver também Cartwright (1983), Earman (1986, p. 80-110), e van Fraassen (1989).

[3] Como exemplo dessa atitude, considere a discussão de Hempel (1965) sobre “suporte para contrafactuais” como um critério para distinguir as leis das generalizações acidentais. Embora Hempel concorde que o suporte para contrafactuais é, por assim dizer, diagnóstico da distinção lei/acidente, ele (1965, p. 339) também sustenta que contrafactuais "apresentam notórias dificuldades filosóficas" e, por isso, não podem ser usados para fornecer qualquer vantagem independente na distinção entre leis e generalizações acidentalmente verdadeiras.

[4] Além do requisito de que as leis devem ser generalizações sem exceção, isso inclui os requisitos de que as leis não contenham termos referentes a objetos ou lugares particulares e o requisito de que as leis devem conter apenas predicados projetáveis no sentido de Goodman (1955).

[5] Para uma avaliação semelhante, consulte Salmon (1989, p. 15).

em uma gama substancial de casos particulares sobre os quais generalizações são leis. Se esse acordo existir, pouco importa para o modelo ND, se formos incapazes de formular critérios completamente gerais que distinguem entre leis e generalizações acidentalmente verdadeiras em todos os casos possíveis. Por exemplo, mesmo sem uma formulação adequada do estatuto de lei, podemos certamente concordar que (2.2.2) é uma lei e (2.2.1) não, e isso é tudo o que precisamos para concluir que (2.2.2) pode figurar nas explicações ND enquanto (2.2.1) não pode.

Infelizmente, no entanto, os assuntos nem sempre são tão simples. Uma questão importante levantada pelo modelo ND diz respeito ao status explicativo das chamadas ciências especiais - biologia, psicologia, economia e assim por diante. Essas ciências estão cheias de generalizações que parecem desempenhar um papel explicativo e, ainda assim, falham em satisfazer muitos dos critérios padrão de estatuto de lei. Por exemplo, embora a lei de Mendel da segregação (M) (que afirma que na reprodução sexuada de organismos, cada uma das duas formas alternativas (alelos) de um gene que especifica uma característica em um locus de um determinado organismo tenha 0,5 probabilidade de terminar em um gameta) seja amplamente utilizada em modelos de biologia evolutiva, ela possui várias exceções, como o impulso meiótico. Um argumento semelhante vale para os princípios da teoria da escolha racional (como a generalização de que as preferências são transitivas) que figuram centralmente na economia. Outras generalizações amplamente utilizadas nas ciências especiais têm escopo muito restrito em comparação com leis paradigmáticas, só se mantêm em regiões espaço-temporais restritas e carecem de integração teórica explícita.

Há uma discordância considerável sobre se essas generalizações são leis. Alguns filósofos (WOODWARD, 2000) sugerem que essas generalizações satisfazem muito poucos critérios padrão para serem consideradas leis, mas podem, no entanto, figurar em explicações; nesse caso, parece que devemos abandonar o requisito ND de que todas as explicações devem apelar para leis. Outros (MITCHELL, 1997), enfatizando diferentes critérios de estatuto de lei, concluem que generalizações como (M) são leis e, portanto, não são ameaças ao requisito de que explicações devem invocar leis. Na ausência de uma formulação de leis mais baseada em princípios, é difícil avaliar essas reivindicações concorrentes e, portanto, é difícil avaliar as implicações do modelo ND para as ciências especiais. De um modo mais geral, na ausência de uma formulação geralmente aceita do estatuto de lei, a lógica do contraste fundamental entre leis e não leis, que está no cerne do que o modelo ND exige, não é clara: é difícil avaliar a alegação de que todas as explicações devem

citar leis, sem uma explicação clara do que é uma lei e do que ela contribui para uma explicação bem-sucedida. No mínimo, fornecer uma formulação desse tipo é um item importante dos negócios inacabados para os defensores do modelo de DN.

2.3 Explicação Indutiva Estatística

O modelo ND visa capturar explicações via dedução de leis determinísticas e isso levanta a questão óbvia do status explicativo de leis estatísticas. Essas leis explicam de alguma forma e, em caso afirmativo, o que elas explicam e em que condições? Hempel (1965) distingue duas variedades de explicação estatística. A primeira delas, explicação *dedutiva-estatística* (DE), envolve a dedução de "uma uniformidade estatística mais estreita" de um conjunto mais geral de premissas, pelo menos uma das quais envolve uma lei estatística mais geral. Como a explicação DE envolve dedução do *explanandum* de uma lei, ela segue o mesmo padrão geral que a explicação ND das regularidades. Entretanto, em adição à explicação DE, Hempel também reconhece um tipo distinto de explicação estatística, que ele chama de explicação *indutiva-estatística* ou IE, envolvendo a subsunção de eventos individuais (como a recuperação de uma pessoa em particular da infecção por estreptococos) sob (o que ele considera como) leis estatísticas (como uma lei que especifica a probabilidade de recuperação, dado que a penicilina foi tomada).

Embora o *explanandum* de uma explicação de ND ou DE possa ser deduzido a partir do *explanans*, não se pode deduzir que algum indivíduo em particular, John Jones, tenha se recuperado a partir da lei estatística acima e da informação de que ele tomou penicilina. No máximo, o que se pode deduzir dessas informações é que a recuperação é mais ou menos provável. Na explicação IE, a relação entre o *explanans* e *explanandum* é, nas palavras de Hempel, "indutiva" ao invés de dedutiva – daí o nome explicação estatística-indutiva. Os detalhes da formulação de Hempel são complexos, mas a ideia fundamental é mais ou menos a seguinte: uma explicação IE será boa ou bem-sucedida na medida em que seu *explanans* confere alta probabilidade ao seu *explanandum* resultante.

Portanto, se é lei estatística que a probabilidade de recuperação do estreptococo, considerando que alguém tomou penicilina, é alta e Jones tomou penicilina e se recuperou, essa informação pode ser usada para fornecer uma explicação IS da recuperação de Jones. No entanto, se a probabilidade de recuperação for baixa (por exemplo, menor que 0,5), considerando que Jones tomou penicilina,

então, mesmo que Jones se recupere, não podemos usar essas informações para fornecer uma explicação IE sobre sua recuperação.

2.4 Motivação para o modelo ND: Expectabilidade Nômica e uma Formulação de Regularidade da Causalidade

Por que supor que todas (ou mesmo algumas) explicações tenham uma estrutura ND ou IE? Há duas ideias que desempenham um papel motivador central na discussão de Hempel (1965). O primeiro conecta as informações fornecidas por um argumento ND com uma certa concepção do que é alcançar o entendimento do porquê qual algo acontece - apela a uma ideia sobre o objeto ou o ponto de dar uma explicação. Hempel escreve:

> [...] uma explicação *ND* responde à pergunta " *Por que* o fenômeno-*explanandum* ocorreu?" mostrando que o fenômeno resultou de certas circunstâncias particulares, especificadas no C_1, C_2, ... C_k, de acordo com as leis L_1, L_2, ..., L_R. Ao apontar isso, o argumento mostra que, dadas as circunstâncias particulares e as leis em questão, era de se esperar a ocorrência do fenômeno; e é nesse sentido que a explicação nos permite *entender por que* o fenômeno ocorreu (1965, p. 337, grifo do autor).

Pode-se pensar na explicação IE como envolvendo uma generalização natural dessa ideia. Embora uma explicação IE não mostre que o fenômeno-*explanandum* era esperado com certeza, ela faz a melhor coisa a seguir: mostra que o fenômeno-*explanandum* é pelo menos esperado com alta probabilidade e, dessa maneira, fornece entendimento. Em termos mais gerais, os modelos ND e IE compartilham a ideia comum de que, como Salmon (1989) coloca, "a essência da explicação científica pode ser descrita como *expectativa nômica*, ou seja, a expectativa com base em conexões com estatuto de lei" (1989, p. 57).

A segunda motivação principal para o modelo ND/IE tem a ver com o papel das alegações causais na explicação científica. Há um desacordo considerável entre os filósofos sobre se todas as explicações na ciência e na vida cotidiana são

causais e, também há desacordo sobre em que consiste a distinção (se houver) entre explicações causais e não causais[6]. Não obstante, praticamente todos, incluindo Hempel, concordam que muitas explicações científicas citam informações sobre causas. No entanto, Hempel, juntamente com a maioria dos outros primeiros defensores do modelo ND, não está disposto a tomar a noção de causalidade como primitiva na teoria da explicação - ou seja, ele não estava disposto a simplesmente dizer que *X* se apresenta em uma explicação de *Y* se e apenas se *X* causa *Y*. Em vez disso, os adeptos do modelo ND geralmente procuravam uma formulação de causalidade que satisfizesse os requisitos empiristas descritos na Seção 1. Em particular, os defensores do modelo ND geralmente aceitam uma teoria de causalidade amplamente humeana ou de regularidade, segundo a qual (muito aproximadamente) todas as reivindicações causais implicam a existência de alguma regularidade correspondente (uma "lei") que vincula causa a efeito. Isso é levado a mostrar que todas as explicações causais "implicam", talvez apenas "implicitamente", que essa lei/regularidade existe e, portanto, que as leis estão "envolvidas" em todas essas explicações, exatamente como o modelo ND afirma.Para ilustrar essa linha de argumento, considere

(2.4.1) O impacto do meu joelho na mesa causou a queda do tinteiro.

---

[6] Por exemplo, de acordo com Salmon (1984), todas as explicações (porque) são causais e as explicações causais requerem o rastreamento de processos causais e suas interseções (ver seção 4). Por outro lado, Graham Nerlich (1979) está de acordo com Salmon sobre o que conta como uma explicação causal, mas afirma que existe uma importante forma não causal de explicação que ele chama de explicação geométrica - por exemplo, a explicação das trajetórias de partículas no campo gravitacional por referência à estrutura afim do espaço-tempo. Salmon presumivelmente negaria que tais apelos à estrutura do espaço-tempo sejam explicativos. Sober (1983) oferece outra distinção entre formas causais e não causais de explicação: ele compara as explicações que traçam a sequência real de eventos que levam a algum resultado que ele pensa como causal, com uma forma de explicação não causal que ele chama *explicação de equilíbrio* em que um resultado é explicado mostrando que um grande número de estados iniciais de um sistema irá evoluir de tal forma que o sistema termina no estado de resultado que desejamos explicar, mas no qual nenhuma tentativa é feita para rastrear a sequência real de eventos que levaram a esse resultado.

(2.4.1) é uma chamada explicação causal singular, apresentada por Michael Scriven (1962) como um contraexemplo à alegação de que o modelo ND descreve as condições necessárias para uma explicação bem-sucedida. De acordo com Scriven, (2.4.1) explica a queda do tinteiro, embora nenhuma lei ou generalização apareça explicitamente em (2.4.1) e (2.4.1) pareça consistir em uma única frase, em vez de um argumento dedutivo. A resposta de Hempel (1965, p. 360ss) é que a ocorrência de "causou" em (2.4.1) não deve ser deixada sem ser analisada ou tomada como explicativa, exatamente como está. Em vez disso (2.4.1) deve ser entendido como alegando "implicitamente" ou "tacitamente" que existe uma "lei" ou regularidade que vincula os impactos do joelho à queda de tinteiros. Segundo Hempel, é a alegação implícita de que alguma dessas leis existem que "distingue" (2.4.1) de "uma mera narrativa sequencial" na qual se diz que o derramamento segue o impacto, mas sem nenhuma reivindicação de conexão causal - uma narrativa que (Hempel pensa) claramente não seria explicativa. Esta lei de ligação é a premissa nomológica no argumento ND que, de acordo com Hempel, é "implicitamente" afirmado por (2.2.1).

Existem duas maneiras relacionadas, mas distintas, de entender esse argumento, ambas sugeridas por partes da discussão de Hempel. De acordo com a primeira, a afirmação de Hempel é que a estrutura subjacente real de (2.4.1) é algo como:

> (2.4.2) *(L)* Sempre que os joelhos impactam nas mesas em que um tinteiro se encontra e outras condições *K* são cumpridas (onde *K* especifica que o impacto é suficientemente forte, etc.), o tinteiro tombará. (Referência a *K* é necessária, pois o impacto dos joelhos na mesa com tinteiros nem sempre resulta em tombamento.)
> *(I)* Meu joelho impactou uma mesa em que um tinteiro se encontra e outras condições *K* são atendidas.
> *(E)* O tinteiro cai.

Portanto, na medida em que seja explicativo, (2.4.1) "implicitamente" satisfaz afinal os requisitos ND/IE, é um argumento ND/IE (ou seja, 2.4.2) disfarçado.

Há uma segunda interpretação do argumento de Hempel que, diferentemente da primeira interpretação, não exige que pensemos no conteúdo completo de (2.4.2) como de alguma forma já implícito em (2.4.1). Em vez disso, (2.4.2) cumpre o papel de um *ideal* contra o qual (2.4.1) deve ser medido. (2.4.2) explicita quais informações uma explicação completa e totalmente adequada para E precisaria conter - informações presentes em (2.4.1) apenas de maneira parcial ou incompleta. Nesta visão do tema, pensamos em (2.4.1) como um esboço de *explicação* (HEMPEL, 1965b, p. 423) que transmite algumas das informações transmitidas por (2.4.2) ou aponta na direção da explicação mais completa (2.4.2). Idealmente, explicações causais singulares como (2.4.1) devem ser substituídas por explicações ND explícitas como (2.4.2).

Em qualquer uma das interpretações, no entanto, a ideia básica é que uma explanação adequada do papel das alegações causais na explicação conduz, por meio de uma teoria de causalidade humeana ou de regularidade, à conclusão de que, pelo menos idealmente, explicações devem satisfazer o modelo ND/IE. Vamos chamar essa linha de argumento de "estrutura oculta" em reconhecimento à função que ela atribui a uma estrutura ND oculta (ou pelo menos não explícita) que é reivindicada como associada com (2.4.1).

Essa estratégia será examinada na seção 2.6, mas deixe-me primeiro comentar sobre uma característica da discussão até agora que pode parecer intrigante. Os limites da categoria "explicação científica" estão longe de serem claros, mas embora (2.4.1) seja sem dúvida uma explicação, não é o que geralmente se pensa como "ciência" - ao contrário, é uma reivindicação da "vida comum" ou "senso comum". Isso levanta a questão de por que os adeptos do modelo ND/IE não respondem simplesmente ao suposto contraexemplo (2.4.1) negando que seja uma instância da categoria "explicação científica" - ou seja, alegando que o modelo ND/IE não é uma tentativa de reconstruir a estrutura de explicações como (2.4.1), mas que, ao invés, destina-se apenas a aplicar-se a explicações que são adequadamente consideradas como "científicas". O fato de que essa resposta nem sempre é adotada pelos defensores do modelo ND é uma indicação da extensão em que, como observado na seção 1, é assumido implicitamente na maioria das discussões sobre explicação científica que existem semelhanças ou continuidades importantes na estrutura entre explicações como (2.4.1) e explicações mais obviamente científicas e que essas semelhanças devem ser capturadas por alguma formulação comum que se aplica a ambas. Na verdade, é uma característica marcante não apenas de Hempel (1965), mas de muitos outros tratamentos de explicação científica, que boa parte da discussão se concentra de fato em explicações causais singulares da “vida comum” semelhantes a (2.4.1),

sendo a suposição tácita que as conclusões sobre a estrutura de tais explicações têm implicações diretas no entendimento da explicação na ciência.

2.5 Entendimento Explicativo e Expectabilidade Nômica: Contraexemplos à Suficiência

Como explicado acima, exemplos como (2.4.1) são possíveis contraexemplos à alegação de que o modelo ND fornece condições necessárias para explicação. Há também vários contraexemplos conhecidos da alegação de que o modelo ND fornece condições **suficientes** para uma explicação científica bem-sucedida. Aqui estão dois exemplos.

Assimetrias explicativas. Existem muitos casos em que a derivação de um *explanandum* E de uma lei **L** e condições iniciais *I* parece explicativo, mas uma derivação "ao contrário" de **I** a partir de **E** e da mesma lei **L** não parece ser explicativa, embora a última, como a primeira, pareça atender aos critérios para uma explicação bem-sucedida ND. Por exemplo, pode-se derivar o comprimento *s* da sombra projetada por um mastro a partir da altura **h** do mastro e do ângulo θ do sol acima do horizonte e leis sobre a propagação retilínea da luz. Essa derivação atende aos critérios ND e parece explicativa. Por outro lado, uma derivação (2.5.1) de **h** a partir de **s** e **θ** e as mesmas leis também atendem aos critérios ND, mas não parecem explicativas. Exemplos como este sugerem que pelo menos algumas explicações possuem características direcionais ou assimétricas às quais o modelo ND é insensível.

Irrelevâncias explicativas. Uma derivação pode satisfazer os critérios ND e, no entanto, ser uma explicação defeituosa, pois contém irrelevâncias além daquelas associadas às características direcionais da explicação. Considere o exemplo de Wesley Salmon (Salmon, 1971, p. 34):

(2.5.2) *(***L***)* Todos os homens que tomam pílulas anticoncepcionais regularmente não engravidam
*(***K***)* John Jones é um homem que toma regularmente pílulas anticoncepcionais
*(***E***)* John Jones não consegue engravidar

É discutível se *(***L***)* atende aos critérios de estatuto de lei impostos por Hempel e muitos outros escritores (se alguém quiser negar que **L** é uma lei, é

preciso alguma base geralmente aceita de princípios para esse julgamento e, como explicado acima, não está claro qual seja essa base). Além disso, (2.5.2) é certamente um argumento dedutivo sólido no qual **L** ocorre como uma premissa essencial. Não obstante, a maioria das pessoas julga que *(***L***)* e **(K***)* não são explicações para **E**. Existem muitas outras exemplificações semelhantes. Por exemplo (Kyburg, 1965), é presumivelmente uma lei (ou pelo menos uma generalização de apoio contrafactual sem exceção) que todas as amostras de sal de mesa que foram enfeitiçadas ao serem tocadas com a varinha de uma bruxa se dissolvam quando colocadas na água. Pode-se usar essa generalização como premissa em uma derivação ND que tem como conclusão de que uma amostra particular de sal enfeitiçado se dissolveu em água. Mas, novamente, o feitiço é irrelevante para a dissolução e essa derivação não é explicação.

Um diagnóstico óbvio das dificuldades colocadas por exemplos como (2.5.1) e (2.5.2) enfoca o papel da causação na explicação. De acordo com esta análise, para explicar um resultado devemos citar suas causas e (2.5.1) e (2.5.2) não conseguem fazer isso. Como Salmon (1989, p.47) coloca, "um mastro de bandeira de uma certa altura causa uma sombra de um determinado comprimento e, portanto, explica o comprimento da sombra". Por outro lado, “a sombra não causa o mastro e, consequentemente, não pode explicar sua altura”. Da mesma forma, tomar pílulas anticoncepcionais não causa a falha de Jones em engravidar e é por isso que (2.5.2) falha em ser uma explicação aceitável. Nesta análise, o que (2.5.1) e (2.5. 2) mostram é que uma derivação pode satisfazer os critérios ND e ainda falhar em identificar as causas de um *explanandum* - quando isso acontece, a derivação deixará de ser explicativa.

Como explicado acima, os defensores do modelo ND não considerariam esse diagnóstico muito esclarecedor, a menos que acompanhado de algum relato de causalidade que não toma simplesmente essa noção como primitiva (na verdade, Salmon fornece essa formulação, que consideraremos na Seção 4). Devemos observar, no entanto, que uma lição aparente de (2.5.1) e (2.5.2) é que a abordagem de regularidade da causação favorecida pelos teóricos do modelo ND é, na melhor das hipóteses, incompleta: a ocorrência de *c*, e a existência de alguma regularidade ou lei que os vincule (ou **x** tendo propriedade **P** e **x** tendo propriedade **Q** e alguma lei que os vincule) não é uma condição suficiente para a verdade da alegação de que **c** causou **e** ou **x** ter **P** é causalmente ou explicativamente relevante para **x** ter **Q**. De maneira mais geral, se os contraexemplos (2.5.1) e (2.5.2) forem aceitos, se segue que o modelo ND falha ao estipular condições suficientes para explicação. Explicar

um resultado não é apenas uma questão de mostrar que é nomicamente esperável.

Existem duas reações possíveis que se pode ter com relação a essa observação. Uma é que a ideia de que a explicação é uma questão de expectativa nômica está correta até certo ponto, mas que algo mais é também exigido. De acordo com essa avaliação, o modelo ND/IE indica uma condição **necessária** para uma explicação bem-sucedida e, além disso, uma condição que não é uma parte redundante em um conjunto de condições que são conjuntamente suficientes para a explicação. No entanto, algum outro recurso independente, **X** (que explicará os recursos direcionais da explicação e garantirá o tipo de relevância explicativa que aparentemente está faltando no exemplo de controle de natalidade) deve ser adicionado ao modelo ND para obter um relato bem-sucedido da explicação. A ideia é que a Expectabilidade Nômica + **X** = Explicação. Algo como essa ideia é endossado pelos modelos unificacionistas de explicação desenvolvidos por Friedman (1974) e Kitcher (1989), discutidos na Seção 5 abaixo.

Uma segunda conclusão possível, mais radical, é que a formulação ND do objetivo ou lógica da explicação está errada de uma maneira muito mais fundamental e que o modelo ND nem mesmo estipula condições necessárias para uma explicação bem-sucedida. Como observado acima, a menos que o argumento da estrutura oculta seja aceito, essa conclusão é fortemente sugerida por exemplos como (2.4.1) ("O impacto do meu joelho causou a queda do tinteiro") que parecem envolver explicação sem a citação explícita de uma lei ou estrutura dedutiva. Para avaliar se o modelo ND/IE fornece condições necessárias para explicação, devemos considerar a estratégia da estrutura oculta em mais detalhes.

2.6 A Estratégia de Estrutura Oculta

Pode parecer que a contenção da estratégia de estrutura oculta de que explicações causais singulares como (2.4.1) são explicações implícitas ND/IE ou esboços de tais explicações seja, na melhor das hipóteses, relevante para a questão de saber se o modelo ND/IE fornece uma reconstrução adequada desse tipo específico de explicação. De fato, no entanto, a estratégia de Hempel de tratar as explicações como dispositivos para transmitir informações, mas de maneira "parcial" ou "incompleta", sobre explicações "ideais" fundamentais de uma forma prima facie bastante diferente que estão pelo menos parcialmente epistemicamente ocultas daqueles que usam a explicação original, não ideal, continuaram a ser muito

populares nas teorizações recentes sobre a explicação científica. Essa estratégia forma a base, por exemplo, do contraste de Peter Railton (1978, 1981) entre um "texto explicativo ideal" que contém todas as informações causais e nomológicas relevantes para algum resultado de interesse e as explicações "não ideais" que realmente damos, como (2.4.1). De acordo com Railton, este último fornece “informações explicativas” em virtude de transmitir informações sobre uma parte ou aspecto limitado do texto ideal e são explicativos em virtude de fazê-lo. A estratégia de estrutura oculta também desempenha um papel importante na formulação unificacionista da explicação desenvolvida por Philip Kitcher (1989), que também insiste que devemos "distinguir entre o que é dito em uma ocasião em que a informação explicativa é dada e a explicação ideal fundamental" (KITCHER, 1989, p. 414). De fato, qualquer teoria da explicação que, como o modelo unificacionista de Kitcher, insiste em que leis (ou generalizações de generalidade considerável) e estrutura dedutiva são condições necessárias para uma explicação bem-sucedida deve apelar para algo como estratégia de estrutura oculta, uma vez que é geralmente aceito existirem muitas explicações aparentes que não estão de acordo com essas condições em sua estrutura manifesta.

Embora a estratégia de estrutura oculta mereça mais atenção do que pode receber aqui, vários pontos parecem claros. Primeiro, a noção de uma explicação "transmitindo informações sobre" outra explicação "fundamental" exige uma discussão considerável. Dependendo do significado de "fundamental", é discutível que existam muitas explicações fundamentando (2.4.1) – (i) a explicação (2.4.2), assumindo que a condição **K** pode ser especificada de maneira não trivial; (ii) uma explicação no nível da física clássica que faça referência a leis que regem colisões inelásticas, o comportamento de líquidos quando não estão confinados a recipientes e assim por diante; e (iii) uma explicação em que o comportamento de todo o sistema é caracterizado em termos de alguma teoria física mais fundamental (mecânica quântica, teoria das supercordas etc.). Estão **todas** estas explicações implícitas em (2.4.1) ou (2.4.1) transmite informações parciais sobre **todas** elas? Em que sentido "implícito" ou "transmite informações sobre" isso poderia ser verdade?

Railton (1981) sugere que uma afirmação explicativa fornece informações sobre um texto ideal fundamental se a primeira reduzir a incerteza sobre algumas das propriedades do texto, no sentido de incluir ou excluir várias possibilidades relacionadas à sua estrutura. Como Railton reconhece, essa proposta tem muitas consequências contraintuitivas. Para usar o próprio exemplo de Railton, “o texto ideal relevante contém mais de 10^2 palavras em Inglês”, se verdadeiro, conta como

explicação para um episódio de decaimento radioativo (1981, p. 246). Da mesma forma, a alegação de que **X** e **Y** estão **correlacionados** contará como uma explicação parcial de **X** e **Y** na suposição plausível de que essa alegação transmite a informação de que uma das três possibilidades provavelmente é verdadeira – ou **X** causa **Y** ou **Y** causa **X** ou eles têm uma causa comum - e, portanto, reduz a incerteza sobre o conteúdo do texto fundamental ideal. Isso contrasta com o amplo juízo de que correlações em si não são explicativas. De fato, em uma visão como a de Railton, mesmo a afirmação de que algum resultado não tem causas ou não é governada por nenhuma lei conta como uma "explicação" desse resultado, supondo que essa afirmação seja verdadeira. De fato, tal afirmação é aparentemente **maximamente** explicativa, uma vez que transmite tudo o que há para ser dito sobre o texto explicativo ideal associado a esse evento. Exemplos como esses sugerem que nem toda afirmação que reduz a incerteza sobre o conteúdo de um texto explicativo ideal deve ser considerada explicativa – tal visão permite que muito conte como explicação.

Seria plausível considerar o texto que contém todos os detalhes completos de informações causais e nomológicas relevantes para algum resultado como pelo menos um “ideal” contra o qual várias explicações candidatas desse resultado devam ser julgadas? Suponha que nos seja apresentada uma explicação da economia ou da psicologia que não apele a nenhuma generalização que estamos preparados para considerar como uma lei, mas que subjacente a essa explicação "não ideal" esteja um conjunto incrivelmente complexo de fatos descritos em termos de mecânica clássica e eletromagnetismo, juntamente com as leis relevantes dessas teorias. Se, como quase certamente será o caso, essa "explicação" fundamental é computacionalmente intratável e cheia de detalhes irrelevantes (consulte a seção 4 para obter mais informações sobre o que isso pode significar), seria possível perguntar em que sentido ela é um ideal contra o qual a explicação original deve ser medida. A explicação econômica será realmente melhor de acordo com a transmissão do máximo de informações possíveis sobre esses detalhes subjacentes?

Por fim, considere a conexão entre explicação e entendimento. Normalmente se pensa em uma explicação como algo que fornece entendimento. De maneira semelhante, parte da tarefa de uma teoria da explicação é identificar os aspectos estruturais das explicações (ou as informações que elas transmitem) em virtude das quais elas fornecem entendimento. Por exemplo, como observado acima, o modelo ND conecta o entendimento com o fornecimento de informações sobre a expectabilidade nômica - a ideia é que entender por que um resultado ocorre é uma questão de se notar que ele era esperado com base em uma lei. O problema que

isso levanta para a estratégia de estrutura oculta é que as informações associadas à estrutura oculta que supostamente sustentam explicações "não ideais", como (2.4.1), são tipicamente desconhecidas ou epistemicamente inacessíveis para quem usa a explicação. É difícil ver como essa estrutura ou informação pode contribuir para o entendimento se estiver epistemicamente oculta dessa maneira. Por exemplo, parece plausível que muitos (se não quase todos) usuários do (2.4.1) (tanto aqueles que podem oferecer uma explicação e os destinatários que a adotam para fornecer entendimento) não tenham conhecimento da estrutura ND subjacente a ela - de fato, é plausível que muitos usuários não tenham a noção de uma lei da natureza e de um argumento dedutivamente válido e, portanto, qualquer noção de que exista **qualquer** argumento ND (desconhecido) subjacente a (2.4.1). Se for esse o caso, como a mera obtenção dessa estrutura ND, independentemente da percepção de alguém de sua existência, pode funcionar de modo a fornecer entendimento quando (2.4.1) é usado? Em vez disso, parece que os recursos de (2.4.1) que lhe conferem importância explicativa - que o tornam uma explicação - devem ser características que podem ser conhecidas, compreendidas ou reconhecidas por quem usa a explicação. Um ponto semelhante será válido para muitas outras explicações candidatas que falhem em conformidade com os requisitos ND, como explicações de ciências como economia e psicologia que parecem não ter leis.

O que podemos concluir dessa discussão sobre a estratégia de estrutura oculta? Se a estratégia falhar, haverá muitas explicações aparentes que não satisfazem as condições necessárias para a explicação imposta pelo modelo ND/IE. Por outro lado, é possível que haja formas de desenvolver a estratégia de estrutura oculta que respondam adequadamente às dificuldades descritas acima. Se isso for correto, a ideia de que os requisitos ND/IE são pelo menos condições necessárias para a explicação ideal pode, apesar de tudo, ser defensável, embora permaneçam os contraexemplos à suficiência do modelo mencionado.

A formulação mais autoritária e abrangente dos modelos ND e IE é provavelmente Hempel (1965b). Isso é reimpresso em Hempel (1965a), juntamente com vários outros artigos que abordam vários aspectos do problema da explicação científica. Além das referências citadas nesta seção, Salmon (1989, pp. 46ss) descreve uma série de contraexemplos bem conhecidos aos modelos ND/IE e discute seu significado.

3. O Modelo RE

3.1 A ideia básica

Grande parte da literatura subsequente sobre explicação foi motivada por tentativas de capturar as características de relevância causal ou explicativa que parecem ter sido deixadas de fora de exemplos como (2.5.1) e (2.5.2), tipicamente dentro das restrições empiristas descritas acima. O modelo de relevância estatística (ou RE) de Wesley Salmon (Salmon, 1971) é uma tentativa muito influente de capturar essas características em termos da noção de relevância estatística ou relações de dependência condicional. Dada alguma classe ou população **A**, um atributo **C** será **estatisticamente relevante** para outro atributo *B* se e apenas se **P**(**B**|**A.C**) ≠ **P**(**A**|**B**) – isto é, se e somente se a probabilidade da condicional de **B** em **A** e **C** é diferente da probabilidade da condicional de **B** em **A** sozinho. A intuição subjacente ao modelo RE é que as propriedades estatisticamente relevantes (ou informações sobre relações estatisticamente relevantes) são explicativas e as propriedades estatisticamente irrelevantes não o são. Em outras palavras, a noção de uma propriedade fazendo diferença para um *explanandum* é especificada em termos de relações de relevância estatística.

Para ilustrar essa ideia, suponha que no exemplo das pílulas anticoncepcionais (2.5.2) a população original **T** inclui ambos os sexos. Então

P(Gravidez | **T**.Homem.Toma pílula anticoncepcional)
= **P**(Gravidez | *T*.Homem)
= 0
enquanto
P(Gravidez | **T**.Mulher.Toma pílula anticoncepcional)
≠ **P**(Gravidez | **T**.Mulher)

assumindo que nem todas as mulheres na população tomam pílulas anticoncepcionais. Em outras palavras, se você é um homem nesta população, tomar pílulas anticoncepcionais é estatisticamente irrelevante para saber se você engravidou, enquanto se você for uma mulher isso é relevante. Dessa forma, podemos captar a ideia de que tomar pílulas anticoncepcionais é explanatoriamente irrelevante para a gravidez entre os homens, mas não entre as mulheres.

Para caracterizar o modelo RE com mais precisão, precisamos da noção de uma *partição homogênea*. Uma partição homogênea de **A** é um conjunto de subclasses ou células $\mathbf{C}_i$ de **A** que são mutuamente exclusivos e exaustivos, onde $\mathbf{P}(\mathbf{B} \mid \mathbf{A}.\mathbf{C}_i) \neq \mathbf{P}(\mathbf{B} \mid \mathbf{A}.\mathbf{C}_j)$ para todos $\mathbf{C}_i \neq \mathbf{C}_j$ e onde nenhuma partição estatisticamente relevante de qualquer uma das células **A**, $\mathbf{C}_i$ pode ser feita em relação a **B**, isto é, não há atributos adicionais $\mathbf{D}_k$ em **A**, de tal modo que $\mathbf{P}(\mathbf{B} \mid \mathbf{A}.\mathbf{C}_i) \neq \mathbf{P}(\mathbf{B} \mid \mathbf{A}.\mathbf{C}_i.\mathbf{D}_k)$.

No modelo SR, uma explicação de o porquê algum membro **x** da classe caracterizada pelo atributo **A** possui o atributo **B** consiste nas seguintes informações:

I. A probabilidade anterior de **B** em **A** : $\mathbf{P}(\mathbf{B} \mid \mathbf{A}) = \mathbf{p}$.
II. Uma partição homogênea de **A** em relação a **B**, $(\mathbf{A}.\mathbf{C}_1,...,\mathbf{A}.\mathbf{C}_n)$, junto com a probabilidade de **B** dentro de cada célula da partição: $\mathbf{P}(\mathbf{B} \mid \mathbf{A}.\mathbf{C}_i) = \mathbf{p}_i$ e
III. A célula da partição na qual **x** pertence.

Para empregar um dos exemplos de Salmon, suponha que queremos construir uma explicação RE de porque **x**, que tem infecção estreptocócica = **S**, recupera rapidamente = **Q**. Seja **T***(***-T***)* consoante a se **x** é (não é) tratado com penicilina, e **R***(***-R***)* = consoante a se o sujeito tem uma cepa resistente à penicilina. Suponha, como hipótese de trabalho, que nenhum outro fator é relevante para a recuperação rápida. Existem quatro combinações possíveis dessas propriedades: **T**.**R**, **-T**.**R**, **T**.**-R**, **-T**.**-R**, mas vamos supor que

$$\begin{aligned}&\mathbf{P}(\mathbf{Q}|\mathbf{S}.\mathbf{T}.\mathbf{R})\\ =\ &\mathbf{P}(\mathbf{Q}|\mathbf{S}.\text{-}\mathbf{T}.\mathbf{R})\\ =\ &\mathbf{P}(\mathbf{Q}|\mathbf{S}.\text{-}\mathbf{T}.\text{-}\mathbf{R})\\ \neq\ &\mathbf{P}(\mathbf{Q}|\mathbf{S}.\mathbf{T}.\text{-}\mathbf{R})\end{aligned}$$

ou seja, a probabilidade de recuperação rápida, dado que se tem estreptococos, é a mesma para quem tem a cepa resistente, independentemente de serem tratados ou não, e também para quem não foi tratado. Em contraste, a probabilidade de recuperação é diferente, presumivelmente maior, entre aqueles com estreptococos que foram tratados e não têm a cepa resistente.

Nesse caso [**S**.(**T**.**R v -T**.**R v -R**.**-T**)], [**S**.**T**.**-R**] é uma partição homogênea de **S** em relação a **Q**. A explicação RE da recuperação de **x** consistirá em um enunciado da probabilidade de recuperação rápida entre todos aqueles com

estreptococos, isto é, (i) acima, um enunciado da probabilidade de recuperação em cada uma das duas células da partição acima, (ii) acima, e a célula na qual **x** pertence, que é **S**.**T**.**R**, (iii) acima. Intuitivamente, a ideia é que essa informação nos diga sobre a relevância de cada uma das combinações possíveis das propriedades **T** e **R** à recuperação rápida entre aqueles com estreptococos e ela é explicativa justamente por esse motivo.

3.2 O modelo RE e Eventos de Baixa Probabilidade

O modelo RE tem uma série de características distintas que geraram uma discussão considerável. Em primeiro lugar, observe que, de acordo com o modelo RE, e em contraste com o modelo ND/IE , uma explicação *não* é um argumento - seja no sentido de um argumento dedutivamente válido em que o *explanandum* segue como uma conclusão do *explanans* ou no sentido de um argumento indutivo em que o *explanandum* segue com alta probabilidade do *explanans*, como no caso da explicação IE. Em vez disso, uma explicação é um conjunto de informações que é estatisticamente relevante para um *explanandum*. Salmon argumenta (e usa o exemplo do controle de natalidade (2.6.2) para ilustrar) que os critérios que um bom argumento deve satisfazer (por exemplo, critérios que asseguram solidez dedutiva ou algum análogo indutivo) são simplesmente diferentes daqueles que uma boa explicação deve satisfazer. Entre outras coisas, como Salmon coloca, "irrelevâncias [são] inofensivas em argumentos, mas fatais em explicações" (1989, p. 102). Conforme explicado acima, ao associar uma explicação bem-sucedida ao fornecimento de informações sobre relações de relevância estatística, o modelo RE tenta acomodar essa observação.

Um segundo ponto intimamente relacionado é que o modelo RE se afasta do modelo IE ao abandonar a ideia de que uma explicação estatística de um resultado deve fornecer informações a partir das quais se segue que o resultado ocorreu com alta probabilidade. Como o leitor pode verificar, a declaração do modelo RE acima não impõe tal requisito de alta probabilidade; em vez disso, mesmo resultados muito improváveis serão explicados, desde que os critérios para a explicação RE sejam atendidos. Suponha que, no exemplo acima, a probabilidade de recuperação rápida de estreptococos, dado o tratamento e a presença de uma cepa não resistente, seja bastante baixa (por exemplo, 0,2). Não obstante, se os critérios (i) a (iii) acima, uma partição homogênea com valores de probabilidade corretos para cada célula

na partição, forem satisfeitos, podemos usar essa informação para explicar por que *x*, que tinha uma cepa não resistente de estreptococos e foi tratado, se recuperou rapidamente. Na verdade, de acordo com o modelo SR, podemos explicar por que algum **x** que é **A** é **B**, mesmo que a probabilidade condicional de **B** dado **A** e a célula $\mathbf{C}_i$ ao qual **x** pertence ($\mathbf{p}_i = \mathbf{P}(\mathbf{B} \mid \mathbf{A}.\mathbf{C}_i)$) é **menor** que a probabilidade prévia (**p** = **P**(**B** | **a**)) de **B** em **A**. Por exemplo, se a probabilidade prévia de recuperação rápida entre todos aqueles com qualquer forma de estreptococos é 0,5 e a probabilidade de recuperação rápida daqueles com uma cepa resistente que não são tratados é 0,1, podemos, no entanto, explicar por que **y**, que atende a essas últimas condições (**-T**.**R**), recuperou-se rapidamente (supondo que sim) citando a célula a que pertence (o fato de ter a cepa resistente e não ter sido tratado), a probabilidade de recuperação dado que ele caiu nesta célula e o outro tipo de informação descrito acima. De modo mais geral, o que importa no modelo RE não é se o valor da probabilidade do resultado do *explanandum* é alto ou baixo (ou mesmo alto ou baixo em comparação com sua probabilidade anterior), mas sim se o suposto explanas cita todos e apenas fatores estatisticamente relevantes e se as probabilidades que invoca estão corretas. Uma consequência disso, que Salmon endossa, embora reconheça que muitos vão considerá-la não intuitiva, é que no modelo RE, o mesmo *explanans* **E** pode explicar tanto um *explanandum* **M** e explananda que são inconsistentes com **M**, tal como **-M**. Por exemplo, o mesmo *explanans* irá explicar por que um sujeito com estreptococos, e certas outras propriedades (por exemplo, **T** e **-R**), recupera-se rapidamente, se o fizer, e também porque ele não se recupera se não o fizer. Por outro lado, nos modelos ND ou IE, se **E** explica **M**, **E** não pode também explicar **-M**.

A intuição de que, ao contrário do modelo IS, o valor que um candidato a *explanans* atribui a um resultado *explanandum* não deve importar para a boa qualidade da explicação que ele fornece pode ser motivada da seguinte maneira. Considere uma moeda genuinamente indeterminista que é fortemente enviesada (p = 0.9) em direção a cara quando lançada. Suponha que, se não for lançada, a moeda tem probabilidade de 0,5 de estar na posição de cara ou coroa e que o fato de a moeda ser lançada ou não é o único fator estatisticamente relevante para se é cara ou coroa. De acordo com o modelo IE, se a moeda for lançada e der cara, podemos explicar esse resultado apelando para o fato de que a moeda foi lançada (uma vez que sob essa condição a probabilidade de cara é alta), mas, se a moeda for lançada e sai coroa, não podemos explicar esse resultado, pois sua probabilidade é baixa. A intuição contrária subjacente ao modelo RE é que entendemos ambos os resultados igualmente bem. O viés da moeda e o fato de a moeda ter sido lançada

são os únicos fatores relevantes para qualquer resultado e esses fatores são comuns a ambos os resultados - uma vez que citamos o lançamento (e especificamos os valores de probabilidade para cara e coroa no lançamento), não deixamos nada de fora que influencie qualquer um dos resultados. Da mesma forma, Salmon argumenta, se for realmente verdade que a partição no exemplo envolvendo recuperação rápida de estreptococos é objetivamente homogênea, se não houver outros fatores que sejam estatisticamente relevantes para a recuperação rápida, além de se o sujeito foi tratado e tem uma cepa resistente, então, uma vez que especificamos a probabilidade de recuperação rápida em todas as combinações desses fatores, e a combinação de fatores possuídos pelo sujeito cuja recuperação (ou não, conforme o caso) que desejamos explicar, especificamos todas as informações relevantes para a recuperação e, nesse sentido, explicamos totalmente o resultado para o assunto[7] .

---

[7] Esse argumento baseia-se na suposição de que, uma vez que alguém tenha especificado todas as informações que são relevantes para algum relato, e não tenha deixado de fora nenhuma informação relevante, ele o terá explicado. Em outras palavras, assume-se que não é possível para todas as informações que são relevantes para algum *M* serrem insuficientes para explicá-lo. Chame isso de argumento Nada Mais a Ser Dito (NMSD). Muitos outros tratamentos de explicação (RAILTON, 1978) fazem uma suposição semelhante. Está longe de ser evidente que essa suposição seja correta. Por que não é possível que, mesmo que dada uma especificação de todas as informações nômicas e causais relevantes para algum explanandum, isso fique aquém do que é necessário para explicá-lo? Por que não pode haver explananda que são simplesmente inexplicáveis? Como uma possibilidade limite, imagine um evento *explanandum* **M**, cuja ocorrência não é governada por leis, que não tem causas, e cuja frequência de ocorrência varia de maneiras imprevisíveis ao longo do tempo e do espaço. Se não há mais nada a ser dito sobre **M** exceto que às vezes ocorre espontaneamente, segue-se que esta informação explica por que **M** ocorre? A intuição contrária é que a ocorrência de **M** é um caso paradigmático de algo que não podemos explicar. Se isso estiver correto, o argumento NMSD não se sustenta em geral e isso levanta a questão de o porquê devemos aceitá-lo no caso particular em que todas as informações relevantes especificam apenas a probabilidade com que um resultado ocorre. Em outras palavras, por que aceitar a afirmação, comum aos modelos IE e RE, de que as teorias estatísticas explicam resultados individuais? Este problema é discutido em mais detalhes abaixo.

3.3 O que Teorias Estatísticas explicam?

Ao avaliar essas afirmações, será útil dar um passo para trás e perguntar do que esses modelos concorrentes de explicação estatística (o modelo IE de Hempel e o modelo RE de Salmon) pretendem ser reconstruções. Na literatura sobre este tópico, duas classes de exemplos ou aplicações aparecem com destaque. Primeiro, há exemplos extraídos da mecânica quântica (MQ). Suponha, por exemplo, que uma partícula tenha uma probabilidade p de penetrar uma barreira potencial que está estritamente entre 0 e 1. Os modelos de explicação estatística presumem que, se a partícula penetrar na barreira, MQ explica esse resultado - os modelos IE e RE pretendem capturar a estrutura de tais explicações. Em segundo lugar, há exemplos extraídos de aplicações biomédicas (ou epidemiológicas) e de ciências sociais, recuperação de estreptococos ou, para citar uma das ilustrações alongadas de Salmon (SALMON, 1971), os fatores relevantes para a delinquência juvenil em meninos adolescentes.

Esta é, para dizer o mínimo, uma classe heterogênea de exemplos. No caso de MQ, o entendimento usual é que os vários resultados de variáveis não ocultas estabelecem que qualquer teoria empiricamente adequada dos fenômenos da mecânica quântica deve ser irredutivelmente indeterminística. É, portanto, plausível que quando usamos a equação de Schrödinger para derivar a probabilidade de que uma partícula com uma certa energia cinética irá atravessar através de uma barreira potencial de uma certa forma, esta representação satisfaça a condição de "homogeneidade objetiva" do modelo RE, não há variáveis adicionais omitidas que afetariam a probabilidade de penetração da barreira. Em contraste, parece bastante improvável que essa condição de homogeneidade seja satisfeita na maioria (na verdade, em qualquer) das ilustrações biomédicas e sociológicas que figuram na literatura sobre explicação estatística. No caso de recuperação de estreptococos, por exemplo, é muito plausível que existam muitos outros fatores além dos dois mencionados acima que afetam a probabilidade de recuperação - esses fatores adicionais incluirão o estado do sistema imunológico do sujeito, várias características do nível geral de saúde do sujeito, o caráter preciso da cepa da doença à qual o sujeito está exposto (resistente *versus* não resistente é quase certamente uma dicotomia de granulação muito grosseira), e assim por diante. Da mesma forma para episódios de delinquência juvenil. Nestes casos, ao contrário dos casos da mecânica quântica, carecemos de uma teoria ou corpo de resultados que delimite os fatores potencialmente relevantes para a probabilidade do resultado que nos

interessa. Assim, em exemplos realistas de conjuntos de fatores estatisticamente relevantes da biomedicina e das ciências sociais, é improvável que a condição de homogeneidade objetiva seja satisfeita ou, em qualquer sentido prático, satisfazível.

Uma diferença relacionada diz respeito à maneira como as evidências estatísticas aparecem nesses dois tipos de aplicações. Alguns fenômenos da mecânica quântica, como a decadência radioativa, são irredutivelmente indeterminísticos. Em contraste, nas aplicações científicas biomédicas e sociais, embora a evidência relevante seja "estatística", normalmente não há uma suposição correspondente de que os fenômenos de interesse são irredutivelmente indeterminísticos. Isso fica particularmente claro em relação aos exemplos das ciências sociais, como fatores de risco para a delinquência juvenil, que Salmon discute. Aqui, a metodologia relevante envolve a chamada modelagem causal ou técnicas de equação estrutural. Pelo menos na maneira mais direta de aplicar tais procedimentos, as equações que determinam se um determinado indivíduo se torna um delinquente juvenil são (se interpretadas literalmente) determinísticas. De acordo com essas abordagens, os fenômenos que estão sendo modelados **parecem** indeterministas porque algumas das variáveis relevantes para o seu comportamento, cuja influência é resumida por um chamado termo de erro, são desconhecidas ou não medidas. As informações estatísticas sobre a incidência de delinquência juvenil entre indivíduos em várias condições desempenham o papel de **evidência** que é usada para estimar parâmetros (os coeficientes) nas equações determinísticas que são tomadas para descrever os processos que regem o início da delinquência. Um ponto semelhante vale para pelo menos muitos exemplos biomédicos[8] .

---

[8] Alguns autores sustentam que a interpretação determinística das equações estruturais é opcional; os termos de erro que figuram em tais equações podem ser interpretados como representando o efeito líquido da operação de processos genuinamente indeterminísticos. Também é discutível que, embora as técnicas convencionais de modelagem causal presumam que as causas macroscópicas da delinquência juvenil são determinísticas, isso não descarta a possibilidade de que existam processos subjacentes em algum nível micro que são indeterministas. Se essas afirmações estiverem corretas, segue-se que as técnicas de modelagem causal são **agnósticas** quanto ao indeterminismo. No entanto, isso não é suficiente para justificar a aplicação do modelo RE para os tipos de exemplos cobertos por técnicas de modelagem causal já que o modelo RE **requer** indeterminismo. Agradecimentos a Elliott Sober pelos comentários úteis sobre este assunto.

Várias conclusões preliminares são sugeridas por essas observações. Em primeiro lugar, está longe de ser óbvio que devamos tentar construir um modelo único e unificado de explicação estatística que se aplique tanto à mecânica quântica quanto a fenômenos macroscópicos como delinquência ou recuperação de uma infecção. Em segundo lugar, e de forma relacionada, embora a explicação em MQ, satisfaça a condição objetiva de homogeneidade, é duvidoso que os tipos de "explicações estatísticas" encontradas nas ciências sociais e biomédicas o façam. Em outras palavras, se uma condição de homogeneidade objetiva é imposta à explicação estatística, não é claro que haverá qualquer exemplo de explicação estatística bem-sucedida fora da mecânica quântica.

Com essas observações em mente, vamos revisitar a questão do que é explicado pelas teorias estatísticas, sejam elas mecânicas quânticas ou macroscópicas. Como vimos, tanto Hempel quanto Salmon, bem como a maioria dos contribuintes posteriores à literatura sobre explicação estatística, tenderam a assumir que teorias estatísticas que atribuem uma probabilidade a algum resultado estritamente entre 0 e 1 devem, não obstante, ser interpretadas como explicando esse resultado. Dado esse ponto de partida comum, Salmon é bastante persuasivo ao argumentar que é arbitrário sustentar, como Hempel faz, que apenas resultados individuais com alta probabilidade possam ser explicados. Mas por que devemos aceitar o ponto de partida? Por que não tomar o argumento de Salmon como uma razão para rejeitar a ideia de que teorias estatísticas explicam resultados individuais, seja de alta ou baixa probabilidade? Se tomarmos essa visão, não precisamos concluir que uma teoria como a MQ é não-explicativa. Em vez disso, podemos considerar o explananda da MQ como fatos sobre as probabilidades ou valores esperados dos resultados, e não como resultados individuais em si. Nesta visão, os explananda que são explicados pela MQ são um subconjunto (apropriado) daqueles que podem ser derivados dele - pelo menos a este respeito, as explicações fornecidas pela MQ são como explicações DE no que concerne à estrutura. Woodward (1989) argumenta que esta interpretação nos permite dizer tudo o que poderíamos legitimamente desejar dizer sobre as virtudes explicativas da MQ. Se isso estiver correto, não há necessidade óbvia de uma teoria separada de explicação estatística de resultados individuais do tipo que Hempel e Salmon procuraram conceber (*vide* nota de rodapé).

No caso da delinquência juvenil e das técnicas de modelagem causal, é, no mínimo, ainda mais intuitivo que o que está sendo explicado não é, por exemplo, porque um determinado menino, Albert, se tornou um delinquente juvenil, mas sim algo mais geral, por exemplo, porque a incidência esperada de delinquência é maior

em certos subgrupos do que em outros. Novamente, tais explananda são dedutíveis do sistema de equações usado para modelar a delinquência juvenil. Ter essa visão do que é explicado pelas teorias estatísticas nos permite evitar várias consequências não intuitivas do modelo de Hempel (por exemplo, que resultados de alta probabilidade são explicados, mas não os de baixa probabilidade) e do modelo de Salmon (por exemplo, o mesmo *explanans* **E** explica ambos **M** e **-M**). No mínimo, aqueles que buscaram construir modelos de explicação estatística de resultados individuais precisam fornecer uma elucidação mais detalhada dos motivos pelos quais tais modelos são necessários e das características de teorização científica que eles devem capturar[9].

---

[9] Correndo o risco de sobrecarregar completamente a paciência dos leitores para este tópico, deixe-me acrescentar que rejeitar as opiniões de Hempel e Salmon sobre os tipos de casos em que é apropriado falar sobre a explicação estatística de resultados individuais e a estrutura de tais explicações não precisa comprometer alguém à posição de que nunca podemos explicar resultados individuais não determinados. Em vez disso, o que se segue é que, se tais explicações de resultados não determinados existem, elas serão diferentes em estrutura dos modelos IE ou RE. Em minha opinião, os contextos indeterminísticos em que é mais natural pensar em termos da explicação dos resultadosindividuais são aqueles em que as causas indeterminísticas são operativas e tais explicações parecem ter uma estrutura bastante diferente daquela que é capturada pelo modelo IE ou RE. Considere a ilustração a seguir da operação de uma causa indeterminística baseada em Dretske e Snyder (1972). Uma fonte radioativa é introduzida em uma câmara por um período de tempo fixo. Se a fonte se deteriorar durante este período, ela acionará um contador Geiger que, por sua vez, liberará um gás venenoso, matando um gato. Suponha que de fato a fonte seja introduzida, ocorra um evento de decomposição e o gato morra. Aqui, parece natural pensar na introdução da fonte como uma causa e como uma contribuição para a explicação da morte do gato, embora a conexão entre esses dois eventos seja indeterminista [você vinha usando "indeterminística"]. Como um insucesso para ambos os modelos IE e RE, considere a seguinte proposta (muito aproximada e esquemática) (WOODWARD, 2003): Uma condição suficiente para a introdução S da fonte para explicar a morte **D** do gato é que as três condições a seguir sejam atendidas. (i) **S** e **D** ocorrem, (ii) se S não houvesse ocorrido, **D** não teria ocorrido, (iii) a probabilidade p de **D** se **S** ocorresse é maior que zero em pelo menos algumas outras ocasiões em que **S** ocorre, onde as condicionais em (ii) e (iii) são interpretadas como contrafactuais sem retrocesso. Chame isso de modelo de causalidade probabilística (**CP**). (A formulação (i) - (iii) se destina a ser aplicada apenas a casos em que há um único caminho causal a partir de **S** para **D** e nenhuma causa preventiva ou potencial de apoio é operativa.) O modelo CP é bastante diferente dos modelos IE e RE. Segundo o modelo CP, em contraste com o modelo IE e de acordo com o

3.4 Relações de Causalidade e Relevância Estatística

Como acabamos de ver, o modelo RE levanta uma série de questões interessantes sobre a explicação estatística de resultados individuais - questões que são importantes independentemente dos detalhes do próprio modelo RE. Esta seção irá abstrair de tais questões e enfocar na motivação primária do modelo RE. Podemos considerar que isso consiste em duas ideias: (i) as explicações devem citar relações causais e (ii) relações causais são capturadas por relações de relevância estatística. Mesmo se (i) for aceito, um problema fundamental com o modelo RE é que (ii) é falso, como um corpo substancial de trabalho[10] , deixou claro, relações causais são muito subdeterminadas pelas relações de relevância estatística. Considere outro exemplo de Salmon (1971): um sistema em que a pressão atmosférica **A** é uma causa comum da ocorrência de uma tempestade **S** e a leitura de um barômetro **B** sem relação causal entre **B** e **S**. Salmon afirma que em tal sistema, **B** e **S** serão correlacionados, mas que **B** é estatisticamente irrelevante para **S** dado **A**, isto é, $P(S \mid A.B) = P(S \mid A)$. Por outro lado, afirma Salmon, **A** permanece relevante para **S** dado **B**, isto é, $P(S \mid A.B) \neq P(S \mid B)$. Similarmente, **S** é irrelevante para **B** dado **A**, mas **A** permanece relevante para **B**, dado **S**. Desta forma, o modelo RE de Salmon tenta capturar a ideia de que **A** é explicativamente, e causalmente, relevante para **S**, enquanto **B** não o é, e que **A** é explicativa e causalmente relevante para **B**, enquanto *S* não o é.

---

modelo RE, o valor de **p** ser alto parece não fazer diferença para o quão boa seja a explicação fornecida. Mesmo que a probabilidade de decomposição fosse muito baixa, se a deterioração ocorrer e o gás for liberado, a introdução da fonte será o que causou e explica a morte do gato. Isso reproduz nosso julgamento pré-analítico sobre o exemplo. Mas, em contraste com o modelo RE, não segue do modelo CP que se **S** ocorre e **-D** ocorre, então a ocorrência de **S** explica **-D**. Isso não acontece, pois o resultado da substituição **-D** para **D** em (ii) é uma afirmação contrafactual que é falsa.

[10] *Vide* Cartwight (1979) e Spirtes, Glymour e Scheines (1993, 2000).

Essas afirmações sobre a conexão entre afirmações causais e relações de relevância estatística são consequências de um princípio mais geral denominado condição Causal de Markov, que foi amplamente discutido na literatura recente sobre causalidade[11] . Um conjunto de variáveis em uma relação causal e uma distribuição de probabilidade associada sobre essas variáveis satisfazem a condição Causal de Markov se, e somente se, condicional às suas causas diretas cada variável é independente de todas as outras variáveis, exceto possivelmente seus efeitos. Dois pontos relevantes emergiram da discussão dessa condição. O primeiro, que foi efetivamente observado pelo próprio Salmon em um trabalho posterior ao seu (1971), é que há circunstâncias em que a condição Causal de Markov falha e, portanto, as alegações causais não implicam nas relações de blindagem descritas acima. Isso pode acontecer, por exemplo, se as variáveis às quais a condição é aplicada forem caracterizadas de forma insuficientemente refinada[12] . O segundo ponto, e uma observação mais fundamental, é que, dependendo dos detalhes do caso, muitos conjuntos diferentes de relações causais podem ser compatíveis com as mesmas relações de relevância estatística, mesmo assumindo que a condição Causal de Markov seja satisfeita. Por exemplo, uma estrutura em que **B** causa **A** que por sua vez causa **S** irá, se assumirmos a condição Causal de Markov (isto é,

---

[11] *Vide*, por exemplo, Spirtes, Glymour and Scheines (1993, 2000), Pearl (2000), Hausman e Woodward (1999).

[12] Um exemplo é fornecido em Salmon (1984) e, subsequentemente, discutido em Spirtes, Glymour e Scheines (1993, 2000). Uma colisão **C** entre uma bola branca e duas outras bolas de bilhar envia a primeira para a caçapa direita (**A**) e a segunda para a esquerda (**B**). **C** é uma causa comum de **A** e **B** e **A** não causa **B**, ou vice-versa. Não obstante, por causa da conservação do momento linear, a informação de que **A** ocorreu fornece informações sobre se **B** ocorreu, mesmo tendo em conta a ocorrência de **C**. Em outras palavras, **A** é estatisticamente relevante para **B**, dado **C**, embora **A** não cause **B**. Intuitivamente, o problema é que a propriedade **C** tem um nível de descrição muito grosseiro. Uma vez que o sistema é presumivelmente determinístico em um nível de descrição mais refinado, no qual especificamos as posições e momentos exatos, **M**, das bolas, condicionar em **M** tornaria **A** independente de **B**, ou seja, irrelevante a **B**. Mas se empregarmos variáveis como **C** que são insuficientemente refinadas, a conexão entre causalidade e relevância estatística assumida, tanto na condição Causal de Markov quanto no modelo **SR**, falhará.

fizermos suposições como as relação de causalidade e relevância estatística de Salmon), conduzirá exatamente às mesmas relações de relevância estatística como no exemplo em que **A** é uma causa comum de **B** e **S**. Da mesma forma se **S** causa **A** que por sua vez causa **B**. Em estruturas com mais variáveis, essa subdeterminação de relações causais por relações de relevância estatística pode ser muito mais extrema. Assim, uma lista de relações de relevância estatística, que é o que o modelo RE fornece, não precisa nos dizer quais relações causais são operativas. Na medida em que essa explicação tem a ver com a identificação das relações causais das quais um resultado explicativo depende, o modelo RE falha em capturá-las totalmente.

Salmon (1971ª) fornece uma apresentação e defesa detalhada do modelo RE. Este ensaio, bem como os artigos de Jeffrey (1969) e Greeno (1970), que defendem visões amplamente semelhantes ao modelo RE, foram coletados em Salmon (1971b). Uma discussão adicional do modelo, bem como uma caracterização mais recente da "homogeneidade objetiva", pode ser encontrada em Salmon (1984). Cartwright (1979) contém algumas críticas influentes ao modelo RE. Teoremas que especificam a extensão precisa da subdeterminação de afirmações causais por evidências sobre relações de relevância estatística podem ser encontrados em Spirtes, Glymour e Scheines (1993, 2000).

4. O Modelo Mecânico-causal

4.1 A ideia básica

Em um trabalho mais recente, Salmon (1984) abandonou a tentativa de caracterizar a explicação ou relações causais em termos puramente estatísticos. Em vez disso, ele desenvolveu uma nova versão que chamou de modelo de explicação Mecânico-causal (MC), uma versão que é semelhante tanto em conteúdo quanto em espírito às chamadas teorias de *processo* de causação do tipo defendido por filósofos como Philip Dowe (2000). Podemos pensar no modelo MC como uma tentativa de capturar o "algo mais" envolvido nas relações causais e explicativas, além dos fatos sobre a relevância estatística, novamente enquanto tentamos permanecer dentro de uma estrutura amplamente Humeana.

O modelo MC emprega várias ideias centrais. Um **processo causal** é um processo físico, como o movimento de uma bola de beisebol no espaço, que se caracteriza pela capacidade de transmitir uma **marca** de forma contínua ("contínua" significa geralmente, embora talvez nem sempre, "espaço-temporalmente contínuo"). Intuitivamente, uma marca é alguma modificação local na estrutura de um processo, por exemplo, um arranhão na superfície de uma bola de beisebol ou um amassado no para-choques do automóvel. Um processo é capaz de transmitir uma marca se, uma vez que a marca é introduzida em uma localização espaço-temporal, ela persistirá em outras localizações espaço-temporais mesmo na ausência de qualquer outra interação. Nesse sentido, a bola de beisebol transmitirá a marca de desgaste de um local para outro. Da mesma forma, um automóvel em movimento é um processo causal porque uma marca na forma de um amassado em um para-choque será transmitida por esse processo de um local espaço-temporal para outro. Processos causais contrastam com **pseudo-processos** que não têm a capacidade de transmitir marcas. Um exemplo é a sombra de um objeto físico em movimento. A ideia intuitiva é que, se tentarmos marcar a sombra modificando sua forma em um ponto (por exemplo, alterando uma fonte de luz ou introduzindo um segundo objeto de oclusão), essa modificação não persistirá a menos que continuamente intervenhamos para mantê-la conforme a sombra ocupa posições espaço-temporais sucessivas. Em outras palavras, a modificação não será transmitida pela própria estrutura da sombra, como seria no caso de um processo causal genuíno.

Devemos observar para referência futura que, conforme caracterizado por Salmon, a capacidade de transmitir uma marca é, em vários sentidos, claramente uma noção contrafactual. Para começar, um processo pode ser um processo causal desde que, mesmo quando de fato não transmita nenhuma marca, seja verdade que, se fosse devidamente marcado, transmitiria a marca. Além disso, a noção de marcação em si envolve um contraste contrafactual - um contraste entre como um processo se comporta quando marcado e como se comportaria se não fosse marcado. Embora Salmon, como Hempel, sempre tenha suspeitado dos contrafactuais, sua opinião na época em que apresentou o modelo MC era que os contrafactuais envolvidos na caracterização da transmissão da marca eram relativamente não problemáticos, em parte porque pareciam experimentalmente testáveis de uma maneira bastante direta. No entanto, a confiança do modelo MC, conforme formulado originalmente, em contrafactuais mostra que ele não satisfaz completamente as restrições Humeanas descritas acima. Em um trabalho subsequente, descrito na Seção 4.4 abaixo, Salmon tentou construir uma versão do modelo MC que evita

completamente a dependência de contrafactuais.

O outro elemento importante no modelo de Salmon é a noção de uma **interação causal**. Uma interação casual envolve uma interseção espaço-temporal entre dois processos causais que modifica a estrutura de ambos - cada processo passa a ter características que não teria na ausência da interação. Uma colisão entre dois carros que amassa os dois é uma interação causal paradigmática.

De acordo com o modelo MC, uma explicação de algum evento **E** irá rastrear os processos causais e interações que levam a **E** (Salmon chama isso de aspecto **etiológico** da explicação), ou pelo menos parte deles, bem como descrever os processos e interações que constituem o próprio evento (o aspecto **constitutivo** da explicação). Desta forma, a explicação mostra como **E** "se encaixa em um nexo causal" (1984, p. 9).

A sugestão de que a explicação envolve "encaixar" um *explanandum* em um nexo causal não nos dá nenhuma caracterização muito precisa de qual a relação entre **E** e outros processos causais e interações deve ser, se as informações sobre o último são para explicar **E**. Não obstante, parece suficientemente claro como a ideia intuitiva deve ser aplicada a exemplos específicos. Suponha que a bola branca, posta em movimento pelo impacto de um taco de bilhar, atinja uma bola oito estacionária com o resultado de que a bola oito é colocada em movimento e a bola branca muda de direção. O impacto do taco também transmite um pouco de giz azul para a bola branca, que é então transferido para a bola oito com o impacto. O taco, a bola branca e a bola oito são processos causais, como é mostrado pela transmissão da marca de giz, e a colisão do taco com a bola branca e a colisão da bola branca e a bola oito são interações causais. A ideia de Salmon é que citar tais fatos sobre processos e interações explica o movimento das bolas após a colisão; por contraste, se uma dessas bolas projeta uma sombra que se move através da outra, isso será causal e explanatoriamente irrelevante para seu movimento subsequente, pois a sombra é um pseudo-processo.

4.2 O Modelo MC e Relevância Explicativa

Como o exemplo da bola branca ilustra, o modelo MC toma como seu paradigma exemplos de interação causal, como colisões em que há "ação por contato" e nenhuma lacuna espaço-temporal na transmissão da influência causal. Há pouca dúvida de que as explicações em que não existem tais lacunas (nenhuma

“ação à distância”) costumam parecer particularmente satisfatórias[13] . No entanto, como Christopher Hitchcock (1995) mostra em um artigo esclarecedor, mesmo aqui o modelo MC deixa de fora algo importante. Considere a "explicação científica" usual e de acordo com as regras básicas elementares do movimento das bolas no exemplo acima após sua colisão. Essa explicação procede derivando esse movimento de informações sobre suas massas e velocidade antes da colisão, a suposição de que a colisão é perfeitamente elástica e a lei da conservação do momento linear. Normalmente pensamos nas informações transmitidas por essa derivação como mostrando que é a massa e a velocidade das bolas, em vez de, digamos, sua cor ou a presença da marca de giz azul, que é explanatoriamente relevante para seu movimento subsequente. No entanto, é difícil ver o que no modelo MC nos permite escolher o momento linear das bolas, em oposição a essas outras características, como explicativamente relevantes. Parte da dificuldade é que, para expressar esses juízos relativamente refinados de relevância explicativa (que é o momento linear em vez de marcas de giz que importa), precisamos falar sobre as relações entre propriedades ou magnitudes e não está claro como expressar tais juízos em termos de fatos sobre processos e interações causais. Tanto o momento linear quanto a marca de giz comunicados à bola branca pelo taco são marcas transmitidas pelo processo causal espaço-temporalmente contínuo que consiste no movimento da bola branca. Ambas as marcas são então transmitidas por meio de uma interação

---

[13] Há evidências da psicologia do desenvolvimento (por exemplo, Leslie e Keeble, 1987) de que mesmo bebês muito jovens são sensíveis à diferença entre as mudanças nas trajetórias das bolas em movimento que envolvem o contato espaço-temporal entre as bolas e as mudanças nas trajetórias que não envolvem tal contato. Em adultos, as primeiras mudanças são muito mais facilmente percebidas (ou julgadas) como envolvendo interações causais do que as últimas. Pode-se conjeturar que os mecanismos psicológicos em ação em tais julgamentos também fundamentam nossa preferência por explicações causais que envolvem ação por contato. Mas, embora as pistas espaço-temporais sejam uma heurística útil para identificar interações causais em alguns casos, como as que envolvem colisões, elas fornecem pouca orientação e, de fato, podem ser positivamente enganosas em outros casos, como veremos a seguir. Portanto, embora a preferência por explicações causais que não contenham lacunas espaço-temporais possa ser muito psicologicamente natural, é duvidoso que forneça uma base útil para a construção de uma teoria geral da explicação causal.

para a bola oito. Parece não haver nada na noção de transmissão de marca ou na noção de um processo causal de Salmon que permite distinguir entre o momento explicitamente relevante e a marca de giz azul explanatoriamente irrelevante.

Ironicamente, como Hitchcock nota, uma observação semelhante pode ser feita sobre o exemplo das pílulas anticoncepcionais (2.5.2) originalmente concebido por Salmon para ilustrar o fracasso do modelo ND para capturar a noção de relevância explicativa. Processos causais espaço-temporais contínuos que transmitem marcas, bem como interações causais, estão em ação quando o RE. Jones ingere pílulas anticoncepcionais, as pílulas se dissolvem, os componentes entram em sua corrente sanguínea, são metabolizados ou processados de alguma forma, e assim por diante. Da mesma forma, processos causais contínuos espaço-temporalmente (embora processos diferentes) estão em ação quando a Sra. Jones toma pílulas anticoncepcionais. No entanto, as pílulas são irrelevantes para a não gravidez do RE. Jones e relevantes para a não gravidez da Sra. Jones. Novamente, parece que a relevância ou irrelevância das pílulas anticoncepcionais para o fracasso do RE. ou Sra. Jones em engravidar não pode ser capturada apenas perguntando se os processos que levaram a esses resultados são processos causais no sentido de Salmon. Um ponto semelhante vale para o exemplo do sal enfeitiçado (2.6.3) - há processos causais espaço-temporalmente contínuos que partem da varinha da bruxa que toca a amostra de sal para os íons individuais **Na** e **Cl** que são formados quando o sal se dissolve, mas isso não é suficiente para que o enfeitiçamento seja causalmente (ou explicativamente) relevante para a dissolução.

Uma maneira mais geral de colocar o problema revelado por esses exemplos é que as características de um processo **P** em virtude do qual se qualifica como um processo causal (capacidade de transmitir marca **M**) podem não ser as características de **P** que são causalmente ou explanatoriamente relevantes para o resultado **E** que queremos explicar (**M** pode ser irrelevante para **E** apesar de alguma outra propriedade **R** de **P** ser a propriedade causalmente relevante para **E**). Portanto, embora a transmissão da marca possa muito bem ser um critério que distingue corretamente entre **processos causais** e **pseudo-processos**, ela não fornece, do modo como está, os recursos para distinguir aquelas **características** ou **propriedades** de um processo causal que são causal ou explicativamente relevantes para um resultado daquelas características que são irrelevantes.

4.3 O Modelo MC e Sistemas Complexos

Um segundo conjunto de preocupações tem a ver com a aplicação do modelo MC a sistemas que se afastam em vários aspectos de paradigmas físicos simples, como a colisão descrita acima. Existem vários exemplos de tais sistemas. Em primeiro lugar, existem teorias como a teoria gravitacional de Newton que envolve "ação à distância" em um sentido fisicamente interessante. Em segundo lugar, há uma série de exemplos da literatura sobre causalidade que não envolvem formas fisicamente interessantes de ação à distância, mas que possivelmente envolvem interações causais sem a intervenção de processos contínuos espaço-temporalmente ou transferência de energia e momentum da causa para o efeito. Isso inclui casos de causalidade por omissão e causalidade por "prevenção dupla" ou "desconexão"[14] . Em todos esses casos, uma aplicação literal do modelo MC parece produzir o juízo de que nenhuma explicação foi fornecida - que a teoria gravitacional newtoniana não é explicativa e assim por diante. Muitos filósofos relutaram em aceitar essa avaliação.

Ainda outra classe de exemplos que levantam problemas para o modelo MC envolve supostas explicações do comportamento de sistemas complexos ou de "nível superior" - explicações que não citam explicitamente processos causais

---

[14] Casos de causalidade por omissão são os casos em que, para colocar a questão de forma intuitiva, a não ocorrência de algum evento causa um resultado, como quando a não prestação de assistência médica por um médico provoca a morte de seu paciente. Nesses casos, não há transferência de energia ou momentum da causa para o efeito e nenhum candidato natural para um processo de conexão. Alguns autores (por exemplo, Dowe, 2000) concluem com base nisso que a causação por omissão não é uma causação real ou literal, embora possua algumas características que a tornam semelhantes aos casos de causação real. Se as omissões podem ser causas ou figurar em explicações causais, isso apresenta um problema prima facie óbvio para teorias de processos causais como a de Salmon. Casos de causalidade por *C* de *E* por dupla prevenção ou desconexão são os casos em que *C* impede ou interfere na operação de um segundo fator *D* que, se operante, bloquearia a ocorrência de *E*. Ao remover o impedidor *D* de *E*, *C* faz com que *E* ocorra. Os exemplos são comuns em contextos biológicos - ver Schaffer (2000), Woodward (2002) e para uma discussão mais geral desse fenômeno, ver Hall (a ser publicado) e Lewis (2000). Novamente, se for aceito que citar uma causa desconectada fornece uma explicação (científica), esta é uma dificuldade para a teoria do processo causal, pelo menos conforme formulada por Salmon.

espaço-temporalmente contínuos envolvendo transferência de energia e momentum, mesmo que possamos pensar que tais processos estão funcionando em um nível mais "fundamental". A maioria das explicações em disciplinas como biologia, psicologia e economia se enquadram nessa descrição, assim como uma série de explicações puramente físicas.

Como ilustração, suponha que um mol de gás está confinado em um recipiente de volume $\mathbf{V_1}$, sob pressão $\mathbf{P_1}$ e temperatura $\mathbf{T_1}$. O gás pode então se expandir isotermicamente em um recipiente maior de volume $\mathbf{V_2}$. Uma maneira padrão de explicar o comportamento do gás - sua taxa de difusão e sua pressão de equilíbrio subsequente $\mathbf{P_2}$, apela às generalizações da termodinâmica fenomenológica, por exemplo, a lei dos gases ideais, a lei da difusão de Graham e assim por diante. Salmon parece considerar supostas explicações baseadas pelo menos na primeira dessas generalizações como não explicativas, porque elas não traçam processos causais contínuos – ele pensa nas moléculas individuais como processos causais, mas não no gás como um todo[15] . No entanto, é absolutamente impossível rastrear os processos causais e as interações representadas por cada um dos 6×10^{23} moléculas que constituem o gás e as sucessivas interações (colisões) que ele sofre com todas as outras moléculas. O tratamento estatístico mecânico usual, que Salmon presumivelmente consideraria explicativo, não tenta fazer isso. Ao invés disso, faz certas suposições gerais sobre a distribuição de velocidades moleculares e as forças envolvidas nas colisões moleculares e, em seguida, usa-as, em conjunto com as leis da mecânica, para derivar e resolver uma equação diferencial (a equação de transporte de Boltzmann) que descreve o comportamento geral do gás. Este tratamento abstrai radicalmente os detalhes dos processos causais envolvendo moléculas individuais particulares e, em vez disso, concentra-se na identificação de variáveis de nível superior que se agregam em muitos processos causais individuais e que figuram em padrões gerais que governam o comportamento do gás.

Este exemplo levanta uma série de questões. O que o modelo MC exige no caso de sistemas complexos nos quais não podemos rastrear processos causais individuais, pelo menos em um nível refinado? Como exatamente o modelo mecânico causal evita a conclusão (desastrosa) de que qualquer explicação bem-sucedida

---

[15] Alguém pode muito bem se perguntar qual é a base para este juízo. Não pode o gás como um todo ser "marcado" - por exemplo, aquecendo-o - e não transmitirá o gás essa marca, pelo menos por algum tempo?

do comportamento do gás deve traçar as trajetórias de moléculas individuais? A explicação mecânica estatística descrita acima rastreia com sucesso processos e interações causais ou especifica um mecanismo causal no sentido exigido pelo modelo MC e, em caso afirmativo, o que exatamente rastrear processos e interações causais envolve ou significa em conexão com tal sistema? Considerando as coisas como estão agora, tanto o modelo MC e as teorias de causalidade do processo, que são seus descendentes mais recentes, estão incompletas.

Há outro aspecto desse exemplo que vale a pena comentar. Mesmo se, *per impossible*, um relato que rastreou trajetórias moleculares individuais fosse produzido, há aspectos importantes em que não forneceria o tipo de explicação do comportamento macroscópico do gás que provavelmente estaríamos procurando - e não apenas porque tal relato seria complexo demais para ser seguido por uma mente humana. Há um grande número de diferentes trajetórias possíveis das moléculas individuais, além das trajetórias realmente tomadas que produziriam o resultado macroscópico, a pressão final $\mathbf{P_2}$, que queremos explicar. Esta informação é certamente explicativamente relevante para o comportamento macroscópico do gás e gostaríamos que nossa explicação acomodasse esse fato. Muito aproximadamente, dadas as leis que regem as colisões moleculares, pode-se mostrar que quase todas (ou seja, todas exceto um conjunto de medida zero) das possíveis posições iniciais e momentos consistentes com o estado macroscópico inicial do gás, caracterizado por $\mathbf{P_1}$, $\mathbf{T_1}$, e $\mathbf{V_1}$, levará a trajetórias moleculares de tal forma que o gás irá evoluir para o resultado macroscópico em que o gás se difunde para um estado de equilíbrio de densidade uniforme através da câmara de pressão $\mathbf{P_2}$. Da mesma forma, há uma grande variedade de diferentes microestados do gás compatíveis com cada um dos vários outros valores possíveis para a temperatura do gás e cada um desses estados levará a uma pressão final diferente $\mathbf{P_2}*$. Se apenas rastrearmos os processos causais (na forma de trajetórias moleculares reais) que levam a $\mathbf{P_2}$, conforme exige o modelo MC, não conseguiremos representar ou capturar essas informações sobre toda a gama de condições sob as quais $\mathbf{P_2}$ e alternativas a ela ocorrerão.

Um ponto semelhante vale para as explicações do comportamento de outros tipos de sistemas complexos, como aqueles estudados em biologia e economia. Considere a explicação padrão, em termos de um deslocamento para cima da curva de oferta, com uma curva de demanda inalterada, para o aumento no preço das laranjas após um congelamento. Subjacente ao comportamento deste mercado estão processos causais e interações individuais espaço-temporalmente contínuas no sentido de Salmon, há uma miríade de transações individuais nas quais o dinheiro

de alguma forma é trocado por bens físicos, as quais envolvem transferências de matéria ou energia, há a troca de informações sobre intenções ou compromissos de comprar ou vender a vários preços, a qual deve ocorrer em algum meio físico e envolver transferências de energia, e assim por diante. Contudo, também parece claro que a produção de uma descrição completa desses processos (supondo, para fins de argumentação, que seja possível fazer isso), produzirá pouco ou nenhum *insight* sobre porque esses sistemas se comportam dessa maneira. Novamente, isso não ocorre apenas porque qualquer "explicação" sobrecarregará nossas habilidades de processamento de informações. É também o caso que grande parte das informações contidas em tal descrição será irrelevante para o comportamento que estamos tentando explicar, pela mesma razão que uma descrição detalhada das trajetórias moleculares individuais conterá informações que são irrelevantes para o comportamento do gás. Por exemplo, enquanto a descrição detalhada dos processos causais individuais envolvidos na operação do mercado de laranjas presumivelmente descreverá se os consumidores individuais compram laranjas em dinheiro, cheque, ou cartão de crédito, se a informação sobre o congelamento for comunicada por telefone ou e-mail, e assim por diante, tudo isso é, em uma primeira aproximação, irrelevante para o preço de equilíbrio, dadas as curvas de oferta e demanda, o preço de equilíbrio será o mesmo enquanto existe um mercado no qual os consumidores podem comprar laranjas por algum meio, estão disponíveis, de alguma forma, informações sobre o congelamento e sobre os preços para compradores e vendedores, e assim por diante[16]. Além disso, aqueles fatores que *são* explanatoriamente relevantes para o preço de equilíbrio, como a forma das curvas de oferta e demanda, não estão, em nenhum sentido óbvio, eles próprios conectados por processos espaço-temporais contínuos ao preço (não está claro até o que essa alegação significa), embora como enfatizado acima, os processos desconhecidos subjacentes à obtenção do equilíbrio são presumivelmente espacial-temporalmente contínuos.

Novamente, a questão é como um relato como o de Salmon pode capturar essa característica de explicação bem-sucedida do comportamento de sistemas complexos - como o relato nos orienta para encontrar o nível “certo” de descrição dos fenômenos

---

[16] O leitor é lembrado aqui de um ponto feito em conexão com a estratégia de estrutura oculta em: as explicações de nível inferior que "sustentam" as explicações de nível superior nem sempre serão "ideais" em termos dos quais as explicações de nível superior devem ser julgadas.

que estamos tentando explicar. Na verdade, como ilustram os exemplos acima, os requisitos que Salmon impõe aos processos causais, e, em particular, o requisito de continuidade espaço-temporal, muitas vezes parecem nos afastar do nível correto de descrição. O nível em que a restrição de continuidade espaço-temporal é mais obviamente respeitada (o nível em que, por exemplo, descrevemos um determinado consumidor como trocando dinheiro por laranjas ou um produtor fazendo um acordo por telefone com um varejista para vender a um determinado preço) parece ser o nível errado para alcançar entendimento.

4.4 Desenvolvimentos mais recentes

Em um trabalho mais recente (SALMON, 1994), impelido em parte por um desejo de evitar certos contraexemplos apresentados por Philip Kitcher (KITCHER, 1989) para sua caracterização da transmissão de marcas, Salmon tentou moldar uma teoria de explicação causal que evita completamente qualquer apelo a contrafactuais. Nesta nova teoria, que é influenciada pela teoria da causalidade de quantidades conservadas de Dowe (2000), Salmon definiu um processo causal como um processo que transmite uma quantidade diferente de zero de uma **quantidade conservada** em cada momento de sua história. Quantidades conservadas são quantidades assim caracterizadas na física, momento linear, momento angular, carga e assim por diante. Uma interação causal é uma interseção de linhas de mundo associadas a processos causais envolvendo troca de uma quantidade conservada. Finalmente, um processo transmite uma quantidade conservada de A para B se possui aquela quantidade em cada estágio sem quaisquer interações que envolvam uma troca dessa quantidade no intervalo semiaberto (**A**,**B**).

Pode-se duvidar de que essa nova teoria realmente evite a dependência de contrafactuais, mas uma dificuldade ainda mais fundamental é que ela ainda não lida adequadamente com o problema da relevância causal ou explicativa descrito acima. Ou seja, ainda enfrentamos o problema de que a característica que torna um processo causal (transmissão de alguma quantidade conservada ou outra) pode nos dizer pouco sobre quais características do processo são causalmente ou explicativamente relevantes para o resultado que queremos explicar. Por exemplo, uma bola de bilhar em movimento transmitirá muitas quantidades conservadas (momento linear, momento angular, carga, etc.) e muitas delas podem ser trocadas durante uma colisão com outra bola. O que é que nos dá o direito de destacar o momento linear das bolas, em vez dessas outras quantidades conservadas como

a propriedade causalmente relevante para seu movimento subsequente? Nos casos em que parece não haver leis de conservação governando a propriedade explicativamente relevante (isto é, casos em que as variáveis explicativamente relevantes não são quantidades conservadas), essa dificuldade parece ainda mais aguda. Propriedades como "ter ingerido pílulas anticoncepcionais", "estar grávida" ou "ser uma amostra de sal enfeitiçado" não figuram nas leis de conservação. Embora se possa dizer que tanto as pílulas anticoncepcionais quanto o sal enfeitiçado são processos causais porque ambos consistem, em algum nível fundamental, de processos que envolvem inequivocamente a transmissão de quantidades conservadas como massa e carga, esta observação por si só não nos diz o que, se é que diz algo, é relevante sobre esses processos fundamentais para a gravidez ou dissolução na água.

Em um artigo ainda mais recente, Salmon (1997) concede esse ponto. Ele concordou que a noção de um processo causal não pode, por si mesma, capturar a noção de relevância causal e explicativa. Ele sugeriu, no entanto, que essa noção pode ser adequadamente capturada apelando para a noção de um processo causal e informações sobre relações de relevância estatística, ou seja, informações sobre relações de (in)dependência condicionais e incondicionais, com a última capturando o elemento de dependência causal ou explicativa que faltava em seu relato anterior:

> Eu diria agora que (1) relações de relevância estatística, na ausência de informações sobre como conectar processos causais, carecem de importância explicativa e que (2) processos causais de conexão, na ausência de relações de relevância estatística, também carecem de importância explicativa (1997, p.476).

Essa sugestão não foi desenvolvida em nenhum detalhe no artigo de Salmon e não é fácil ver como ela pode funcionar. Observamos acima que as relações de relevância estatística muitas vezes subdeterminam grandemente as relações causais entre um conjunto de variáveis. Que razão há para supor que apelar para a noção de um processo causal, no sentido de Salmon, sempre ou mesmo usualmente removerá essa indeterminação? Também observamos que a noção de um processo causal não pode capturar noções refinadas de relevância entre propriedades, que pode haver relevância causal entre instâncias de propriedades das quais (pelo menos no nível de descrição em que são caracterizadas) não estão ligadas por contínuos espaço-temporais ou transferência de quantidades conservadas,

e que as propriedades podem ser vinculadas sem serem causalmente relevantes (lembre-se da marca de giz que é transmitida de uma bola de bilhar para outra). Contanto que seja possível (e por que não deveria ser?) para diferentes afirmações causais implicarem nos mesmos fatos sobre as relações de relevância estatística e para essas afirmações diferirem de maneiras que não podem ser totalmente retiradas em termos das noções de Salmon de processos causais e interações, esta nova proposta também falhará.

Salmon (1984) fornece uma declaração detalhada do modelo Mecânico Causal, conforme formulado originalmente. Salmon (1994, 1997) fornece uma reformulação do modelo e respondem às críticas. Para discussão e crítica do modelo MC, consulte Kitcher (1989, p, 461), Woodward (1989) e Hitchcock (1995).

5. Uma Abordagem Unificacionista de Explicação

5.1 A ideia básica

A ideia básica da teoria **unificacionista** é que explicação científica é uma questão de fornecer uma descrição unificada de uma gama de fenômenos diferentes. Esta ideia é inquestionavelmente atraente. A unificação bem-sucedida pode exibir conexões ou relações entre fenômenos anteriormente considerados não relacionados e isso parece ser algo que esperamos que boas explicações façam. Além disso, a unificação de teorias claramente desempenhou um papel importante na ciência. Exemplos paradigmáticos incluem a unificação de Newton das teorias do movimento terrestre e celestial e a unificação de Maxwell da eletricidade e do magnetismo. A questão-chave, entretanto, é se nossa noção (ou noções) intuitiva de unificação pode ser tornada mais precisa de uma forma que nos permita recuperar as características que pensamos que boas explicações devam possuir.

Michael Friedman (1974) é uma tentativa inicial importante de fazer isso. A formulação de Friedman da ideia unificacionista mostrou sofrer de vários problemas técnicos (KITCHER, 1976) e o desenvolvimento subsequente do tratamento unificacionista da explicação tem sido mais intimamente associado a Philip Kitcher (1989).

Vamos começar apresentando um pouco do vocabulário técnico de Kitcher. Uma **frase esquemática** é uma frase em que parte do vocabulário não lógico foi substituído por letras fictícias. Para usar os exemplos de Kitcher, a frase “Organismos

homozigotos para o alelo falciforme desenvolvem anemia falciforme" está associada a uma série de frases esquemáticas, incluindo "Organismos homozigotos para **A** desenvolvem **P**" e "Para todo **X** se **X** é **O** e **A** então **X** é **P**". As **instruções de preenchimento** são instruções que especificam como preencher as letras fictícias em frases esquemáticas. Por exemplo, as instruções de preenchimento podem nos dizer para substituir **A** com o nome de um alelo e **P** com o nome de um traço fenotípico na primeira das frases esquemáticas acima. **Argumentos esquemáticos** são sequências de frases esquemáticas. As classificações descrevem quais frases em argumentos esquemáticos são premissas e conclusões e quais regras de inferência são usadas. Um **esquema argumentativo** é um triplo ordenado que consiste em um argumento esquemático, um conjunto de conjuntos de instruções de preenchimento, um para cada termo do argumento esquemático, e uma classificação do argumento esquemático. Quanto mais restrições um padrão argumentativo impõe aos argumentos que o instanciam, mais **rigoroso** ele é considerado.

Grosso modo, a ideia norteadora de Kitcher é que explicação é uma questão de derivar descrições de muitos fenômenos diferentes, usando repetidamente o mínimo possível de padrões argumentativos rigorosos - quanto menos forem os padrões usados, mais rigorosos eles são, e quanto maior a gama de diferentes conclusões derivadas, mais unificadas são nossas explicações. Kitcher resume essa visão da seguinte maneira:

> A ciência avança nossa compreensão da natureza, mostrando-nos como derivar descrições de muitos fenômenos, usando o mesmo padrão de derivação repetidamente e, ao demonstrar isso, nos ensina como reduzir o número de fatos que temos de aceitar como definitivos (p. 432).

Kitcher não propõe uma teoria completamente geral de como as várias considerações que ele descreve - número de conclusões, número de padrões e rigor dos padrões – devem ser negociadas umas contra as outras, mas sugere que muitas vezes ficará suficientemente claro o que essas considerações implicam sobre a avaliação de explicações candidatas particulares. Sua estratégia básica é tentar mostrar que as derivações que consideramos como explicações boas ou aceitáveis são instâncias de padrões que, em conjunto, obtêm melhores resultados de acordo com os critérios que acabamos de descrever do que os padrões instanciados

pelas derivações que consideramos explicações defeituosas. Seguindo Kitcher, vamos definir o **depósito explicativo E**(**K**) como o conjunto de argumentos-padrão que unifica ao máximo **K**, o conjunto de crenças aceitas em um determinado momento da ciência. Mostrar que uma determinada derivação é uma explicação boa ou aceitável é, então, uma questão de mostrar que ela pertence ao depósito explicativo.

5.2 Ilustrações do Modelo Unificacionista

Como ilustração, considere a abordagem de Kitcher do problema das assimetrias explicativas (lembre-se da Seção 2.5). Nossas práticas explicativas atuais, chame-as **P**, estão comprometidos com a ideia de que as derivações da altura de um mastro a partir do comprimento de sua sombra não são explicativas. Kitcher compara **P** com uma sistematização alternativa em que tais derivações são consideradas explicativas. De acordo com Kitcher, **P** inclui o uso de um único padrão de explicação de "origem e desenvolvimento" (OD), de acordo com o qual as dimensões dos objetos-artefatos, montanhas, estrelas, organismos etc. são rastreados até "as condições sob as quais o objeto se originou e as modificações que ele subsequentemente sofreu" (1989, p. 485). Agora considere as consequências de adicionar a **P** um padrão adicional **S** (o padrão de sombra) que permite a derivação das dimensões dos objetos a partir de fatos sobre suas sombras. Uma vez que o padrão OD já permite a derivação de todos os fatos sobre as dimensões dos objetos, a adição do padrão de sombra **S** para **P** aumentará o número de padrões argumentativos em *P* e não nos permitirá tirar novas conclusões. Por outro lado, se abandonássemos OD de **P** e o substituíssemos pelo padrão de sombra, não teríamos nenhuma mudança líquida no número de padrões em **P**, mas seríamos capaz de derivar muito menos conclusões do que faríamos com OD, uma vez que muitos objetos não têm sombras (ou sombras suficientes) para derivar todas as suas dimensões. Assim, OD pertence ao depósito explicativo, e o padrão de sombra, não.

A abordagem de Kitcher de outros casos de problemas familiares é semelhante. Por exemplo, ele observa que acreditamos que uma explicação de porque alguma amostra de sal se dissolve na água que apele ao fato de que o sal é enfeitiçado e à generalização (H) de que todo sal enfeitiçado se dissolve na água é defeituosa, pelo menos em comparação com a explicação padrão que apela apenas para a generalização de que (D) todo o sal se dissolve na água. Ele sugere que a "base para essa crença" é que a derivação que apela para (H) instancia um padrão argumentativo que pertence

a uma totalidade de padrões que é menos unificadora do que a totalidade contendo a derivação que apela a (D). Em particular, um depósito explicativo contendo (H) mas não (D) terá um conjunto de consequências mais restrito do que um depósito contendo (D) mas não (H), uma vez que o último, mas não o primeiro, permite a derivação de fatos sobre a dissolução do sal não enfeitiçado na água. E a adição de (H) para um depósito explicativo contendo (D) aumentará o número de padrões sem qualquer ganho de compensação no que pode ser derivado.

Kitcher reconhece que não há nada no relato unificacionista *per se* que exija que toda explicação seja dedutiva: “não há barreira, em princípio, para o uso de argumentos não dedutivos na sistematização de nossas crenças”. No entanto, "a tarefa de comparar o poder unificador de diferentes sistematizações parece ainda mais formidável se os argumentos não dedutivos forem considerados" e, em parte por essa razão, Kitcher endossa a visão de que "em certo sentido, toda explicação é dedutiva." (p. 448).

Qual é o papel da causalidade nesta teoria? Kitcher afirma que "o 'porquê' da causalidade é sempre derivado do 'porquê' da explicação." (1989, p. 477). Isto é, nossos julgamentos causais simplesmente refletem as relações explicativas que surgem de nossas (ou de nossos ancestrais intelectuais) tentativas de construir teorias unificadas da natureza. Não existe uma ordem causal independente além disso que nossas explicações devem capturar. Como muitos outros filósofos, Kitcher leva muito a sério, mesmo se no final ele talvez não endosse totalmente, as preocupações do empirista padrão ou de Hume sobre a acessibilidade epistêmica e a inteligibilidade das afirmações causais. É problemático considerar as noções causais, contrafactuais ou outras pertencentes à mesma família como primitivas na teoria da explicação. Kitcher acredita que é uma virtude de sua teoria não fazer isso, ao invés disso, o autor propõe começar com a noção de unificação explicativa, caracterizada em termos de restrições nas sistematizações dedutivas, onde essas restrições podem ser especificadas de uma forma bastante geral que é independente de noções causais ou contrafactuais, e então mostrar como as reivindicações causais que aceitamos derivam de nossos esforços de unificação.

5.3 As ilustrações criticadas

Conforme observado no início desta seção, a ideia de que a explicação está conectada de alguma forma à unificação é intuitivamente atraente. No entanto,

a maneira particular de Kitcher de sacar essa conexão parece problemática. Considere a abordagem de Kitcher do exemplo do mastro da bandeira. Isso depende muito da verdade contingente de que alguns objetos não projetam sombras suficientes para recuperar todas as suas dimensões. Mas parece fazer parte não apenas do bom senso, mas da teoria física atualmente aceita, que seria impróprio apelar a fatos sobre as sombras lançadas por objetos para explicar suas dimensões, mesmo em um mundo em que todos os objetos lançam sombras suficientes para que todos suas dimensões pudessem ser recuperadas. Não está claro como o relato de Kitcher pode recuperar esse juízo.

A questão fica mais clara se voltarmos nossa atenção para um exemplo variante no qual, ao contrário do exemplo da sombra, existem claramente tantas derivações reversas de efeitos para causas quanto derivações de causas para efeitos. Considere, seguindo Barnes (1992), uma teoria simétrica no tempo como a mecânica newtoniana, aplicada a um sistema fechado como o sistema solar. Chame de **preditivas** derivações do estado de movimento dos planetas em algum momento futuro t a partir de informações sobre suas posições atuais (no momento $\mathbf{t_0}$), massas e velocidades, as forças incidentes sobre eles em $\mathbf{t_0}$, e as leis da mecânica. Agora, compare essas derivações com as derivações **retroativas** em que os movimentos presentes dos planetas são derivados de informações sobre suas velocidades e posições futuras em t, as forças operativas em **t**, e, assim por diante. Parece que haverá tantas derivações retroativas quanto derivações preditivas, e cada uma exigirá premissas exatamente do mesmo tipo geral, informações sobre posições, velocidades, massas etc., e as mesmas leis. Assim, o padrão ou padrões instanciados pelas derivações retroativas parecem exatamente tão unificados quanto o padrão ou padrões associados às derivações preditivas. No entanto, normalmente pensamos nas derivações preditivas e não nas derivações retroativas como explicativas e o estado presente dos planetas como a causa de seu estado futuro e não vice-versa. Mais uma vez, está longe de ser óbvio como as considerações relacionadas com a unificação poderiam gerar tal assimetria explicativa.

Uma resposta possível a este segundo exemplo é aguentar a barra e argumentar que, do ponto de vista da física fundamental, realmente não há diferença na importância explicativa das derivações retroativas e preditivas, e que é uma virtude, não um defeito, da abordagem unificacionista que ela reproduz este juízo. O que quer que se diga a favor dessa resposta, ela não é de Kitcher. Sua afirmação é que nossos julgamentos comuns sobre assimetrias causais podem ser derivados da explicação unificacionista. O exemplo que acabamos de descrever lança dúvidas

sobre essa afirmação. De forma mais geral, lança dúvidas sobre a afirmação de Kitcher de que se pode começar com a noção de unificação explicativa, entendida de uma forma que não pressupõe noções causais, e usá-la para derivar o conteúdo dos julgamentos causais.

5.4 A Heterogeneidade da Unificação

Essa conclusão é reforçada por uma consideração mais geral: a unificação, como aparece na ciência, é uma noção bastante heterogênea, abrangendo muitos tipos diferentes de realizações[17] . Alguns tipos de unificação consistem na criação de um esquema classificatório comum ou vocabulário descritivo onde nenhum esquema satisfatório existia anteriormente, como quando os primeiros investigadores como Lineu construíram sistemas abrangentes e baseados em princípios de classificação biológica. Outro tipo de unificação envolve a criação de uma estrutura matemática comum ou formalismo que pode ser aplicada a muitos tipos diferentes de fenômenos, como quando os sistemas de equações concebidos por Lagrange e Hamilton foram desenvolvidos pela primeira vez em conexão com a mecânica e, em seguida, aplicados a domínios como o eletromagnetismo e termodinâmica. Ainda outros casos envolvem o que pode ser descrito como unificação física genuína, onde fenômenos anteriormente considerados como tendo causas ou explicações bastante diferentes são mostrados como sendo o resultado de um conjunto comum de mecanismos ou relações causais. A demonstração de Newton de que as órbitas dos planetas e o comportamento dos objetos terrestres que caem livremente perto da superfície da Terra são devidas à mesma força da gravidade e obedecem às mesmas leis do movimento foi uma unificação física neste sentido.

---

[17] Morrison (2000) enfatiza a heterogeneidade da unificação e fornece vários exemplos que ilustram o ponto de que muitos tipos de unificação parecem ter pouco a ver com explicação. Sober (1999, no prelo) argumenta que, pelo menos em muitos casos, preferimos teorias unificadas a concorrentes menos unificadas porque as primeiras são mais bem confirmadas do que as últimas, e não porque as primeiras fornecem explicações melhores do que as últimas. Em outras palavras, a unificação é uma virtude epistemológica que é relevante para a confirmação, mas não é uma virtude explicativa.

Destes três tipos de atividades, apenas o terceiro - unificação física - parece ter intuitivamente muito a ver com explicação, pelo menos se pensarmos que a explicação envolve a citação de relações causais. Em particular, dependendo dos detalhes do caso, o tipo de unificação associada à adoção de um esquema classificatório pode nos dizer pouco sobre relações causais. Além disso, como os estudos históricos deixaram claro, um ponto semelhante vale para a unificação formal ou matemática: o fato de que podemos construir uma estrutura matemática comum para lidar com uma gama de fenômenos diferentes de forma nenhuma garante, automaticamente, que identificamos algum conjunto de fatores causais comuns responsáveis por esses fenômenos, isto é, que produzimos uma explicação física unificada deles. Por exemplo, o simples fato de podermos descrever o comportamento de um sistema de massas gravitacionais e a operação de um circuito elétrico por meio das equações de Lagrange não significa que tenhamos alcançado uma explicação comum do comportamento de ambos ou que tenhamos "unificado" a gravitação e a eletricidade em qualquer sentido fisicamente interessante.

Essas considerações levantam a seguinte questão: A versão de Kitcher da unificação é suficientemente discriminativa ou sutil para distinguir aquelas unificações que têm a ver com a explicação de outros tipos de unificação? A preocupação é que ela não seja. A concepção de unificação subjacente à explicação de Kitcher parece ser, no fundo, uma de economia descritiva ou compressão de informação, derivando tanto quanto possível o máximo do mínimo de padrões de inferência. Muitos casos de unificação classificatória e puramente formal envolvendo uma estrutura matemática comum parecem se enquadrar nessa caracterização. Considere esquemas de classificação biológica e esquemas de classificação de objetos geológicos e astronômicos, como rochas e estrelas. Se eu sei que os indivíduos pertencem a uma determinada categoria classificatória (por exemplo, **X** são mamíferos ou ursos polares), posso usar essas informações para derivar muitas de suas outras propriedades (**X** têm coluna vertebral, coração, seus filhotes nascem vivos, etc.) e este é um padrão de inferência que pode ser usado repetidamente para muitos tipos diferentes de **X**. Mas, apesar da disposição de alguns filósofos em considerar tais derivações como explicativas, é prática científica comum considerar tais esquemas como "meramente descritivos" e nos dizendo pouco ou nada sobre

as causas ou mecanismos que explicam por que **X** têm coluna vertebral ou coração[18] .

Outra ilustração do mesmo ponto geral é fornecida pelos numerosos procedimentos estatísticos (análise fatorial, análise de cluster, técnicas de escalonamento multidimensional) que permitem resumir ou representar grandes corpos de informação estatística de uma forma econômica e unificada e derivar fatos estatísticos mais específicos a partir de um conjunto muito menor de suposições pelo uso repetido do mesmo padrão de argumento. Por exemplo, sabendo o "carregamento" de cada um dos **n** testes de inteligência em um único fator comum *g*, pode-se derivar um número muito maior (**n**(**n**-1)/2) de conclusões sobre correlações de pares entre esses testes. Novamente, no entanto, é duvidoso que por si só essa "unificação" nos diga algo sobre as causas do desempenho nesses testes.

5.5 A Concepção O-Vencedor-Leva-Tudo de Unificação Explicativa

Outra dificuldade fundamental com a explicação unificacionista deriva de sua confiança no que pode ser chamado de concepção de unificação do tipo "o vencedor leva tudo". Por um lado, parece que qualquer versão plausível desse relato deve levar à conclusão de que generalizações e teorias podem às vezes ser explicativas com respeito a algum conjunto de fenômenos, embora explicações mais unificadoras desses fenômenos sejam conhecidas[19] . Por exemplo, a lei de

---

[18] Essas observações encobrem algumas questões complexas. Alguns esquemas classificatórios usam informações causais ou etiológicas como base para a construção de classificações. Pode-se argumentar que tais esquemas classificatórios não são meramente descritivos, mas sim explicativos em virtude de "invocar" ou "apelar para" informações causais. Parece claro, entretanto, que existem outros esquemas classificatórios que não são guiados por informações causais, mas, mesmo assim, conseguem compactar informações. Essas classificações não fornecem explicações causais.

[19] Sober (1999) defende a tese mais forte, mas relacionada, de que "não há razão objetiva" (p. 551) para preferir explicações unificadas a desunificadas - o que preferimos dependerá de nossos interesses explicativos. Ele também observa que as explicações que citam uma grande quantidade de micro detalhes e que são consideradas por filósofos como Kitcher como "menos unificadas" do que as explicações que se abstraem de tal detalhe (e, portanto, podem ser aplicadas a sistemas com microestruturas diferentes) não precisam ser

Galileu pode ser usada para explicar fatos sobre o comportamento de corpos em queda, embora forneça uma explicação menos unificadora do que as leis da mecânica newtoniana e da teoria gravitacional, as últimas são, por sua vez, explicativas, embora as explicações que fornecem sejam menos unificadas do que aquelas fornecidas pela Relatividade Geral, as teorias de Coulomb e Ampère são explicativas, embora as explicações que fornecem sejam menos unificadas do que as fornecidas pela teoria de Maxwell, e assim por diante. Se rejeitarmos essa ideia, devemos adotar a conclusão de que, em qualquer domínio, apenas a teoria mais unificada que é conhecida é explicativa; tudo o mais não é explicativo. Chame isso de concepção de o-vencedor-leva-tudo de unificação explicativa.

A concepção do vencedor-leva-tudo desiste da ideia aparentemente muito natural, que se pensaria que o unificacionista gostaria de endossar, que uma explicação pode fornecer menos unificação do que alguma alternativa e, portanto, ser menos profunda ou menos boa, mas ainda se qualificar como explicativa em algum grau. No entanto, a abordagem de Kitcher dos problemas de irrelevância explicativa e assimetria explicativa parece requerer exatamente essa concepção. Por que não podemos apelar para o fato de que essa amostra específica de sal foi enfeitiçada para explicar por que se dissolve? De acordo com Kitcher, qualquer depósito explicativo contendo uma generalização sobre a dissolução do sal enfeitiçado será "menos unificado" do que um depósito explicativo concorrente, de acordo com o qual a dissolução do sal é explicada pelo apelo à generalização de que todo sal se dissolve na água. Da mesma forma, a razão pela qual não podemos explicar a altura de um mastro em termos do comprimento de sua sombra é que as explicações de comprimentos de objetos em termos de fatos sobre sombras não pertencem ao "conjunto de explicações" que "coletivamente fornece a melhor sistematização de nossas crenças" (1989, p. 430). Esta análise requer claramente a ideia do vencedor-

---

consideradas como concorrentes, ou seja, não precisamos escolher entre eles. Uma olhada em qualquer livro-texto de biologia molecular mostrará que Sober está correto ao afirmar que os cientistas nem sempre preferem explicações que se afastem dos microdetalhes e que devemos rejeitar qualquer teoria de explicação que automaticamente requeira tal abstração. É claro que é uma questão adicional se as explicações abstratas deveriam ser consideradas "mais unificadas" do que as explicações que fornecem micro detalhes. Na verdade, as últimas são tipicamente mais integradas com as teorias fundamentais da física e da química e, a esse respeito, indiscutivelmente mais unificadas.

leva-tudo de que uma explicação $\mathbf{T_1}$ que é menos satisfatória do ponto de vista da unificação do que alguma alternativa concorrente $\mathbf{T_2}$ não é explicativa, ao invés de meramente menos explicativa do que $\mathbf{T_2}$. Se Kitcher rejeitasse a ideia do vencedor leva tudo e mantivesse que mesmo que $\mathbf{T_2}$ seja mais unificada do que $\mathbf{T_1}$, disso não segue automaticamente que $\mathbf{T_1}$ não é explicativa, então sua solução para os problemas de irrelevância explicativa e assimetria não estaria mais disponível: sua conclusão deveria ser que uma "explicação" do fracasso do RE. Jones em engravidar em termos de ingestão de pílulas anticoncepcionais é genuinamente explicativa, embora menos do que a explicação alternativa que invoca seu gênero, e da mesma forma para uma derivação da altura de um mastro a partir do comprimento de sua sombra.

Intuitivamente, o problema é que precisamos de uma teoria da explicação que capture várias possibilidades diferentes. Por um lado, existem generalizações e supostas explicações associadas (como a generalização relacionando a pressão barométrica à ocorrência de tempestades e a generalização relacionando o feitiço do sal à sua dissolução na água) que não são de forma alguma explicativa; elas estão aquém do limiar de explicabilidade. Por outro lado, acima desse limiar existe algo mais parecido com um **contínuo**: uma generalização pode ser explicativa, mas fornecer explicações menos profundas ou boas do que alguma alternativa. O que acabamos de ver é que a explicação unificacionista tem dificuldade em capturar simultaneamente essas duas possibilidades. Ou não há limiar (toda derivação é explicativa até certo ponto e é apenas o caso que algumas derivações pertencem a sistematizações que são menos unificadoras e, portanto, menos explicativas do que outras) ou então não há contínuo (apenas as sistematizações mais unificadoras são explicativas).

5.6 A Epistemologia da Unificação

Lembre-se de que, de acordo com Kitcher, o conhecimento causal deriva de nossos esforços de unificação. No entanto, como Kitcher também reconhece, é altamente implausível que a maioria dos indivíduos, deliberada e autoconscientemente, passe pelo processo de comparação de sistematizações dedutivas concorrentes com respeito ao número e rigor de padrões e número de conclusões a fim de determinar qual é mais unificadora. Sua resposta a essa observação é sustentar que a maioria das pessoas adquire conhecimento causal ao absorver o “conhecimento tradicional" de suas comunidades, onde esse conhecimento reflete esforços

sistemáticos anteriores de unificação. Ele escreve que "nosso conhecimento causal cotidiano é baseado em nossa absorção precoce da imagem teórica do mundo que nos foi legada por nossa tradição científica" (1989, p. 469).

Como exatamente essa sugestão deve funcionar? Embora seja certamente verdade que os seres humanos individuais adquiram uma quantidade substancial de conhecimento causal por transmissão cultural, também é óbvio que nem todo conhecimento causal é adquirido dessa maneira. Algum conhecimento causal que os indivíduos adquirem envolve aprender com a experiência. Além disso, a menos que estejamos dispostos a fazer suposições extremamente implausíveis sobre o caráter inato de um grande número de crenças causais específicas, o estoque de conhecimento causal transmitido socialmente deve ter sido adquirido inicialmente de uma forma na qual aprender com a experiência desempenhou um papel importante. A questão que então surge é como esse processo de aprendizagem com a experiência deve funcionar em uma visão como a de Kitcher sobre a fonte de nosso conhecimento causal. Se, como afirma Kitcher, “a ideia de que qualquer indivíduo justifica os juízos causais que faz ao reconhecer os padrões de argumento que melhor unificam suas crenças é claramente absurda” (1989, p. 436), o que é que está acontecendo no nível individual quando as pessoas aprendem com a experiência? Uma possibilidade é que, embora os indivíduos não passem conscientemente pelo processo de comparação do grau de unificação alcançado por sistematizações alternativas quando adquirem novo conhecimento causal aprendendo com a experiência, eles passam por esse processo tácita ou inconscientemente, talvez por causa de alguma disposição geral da mente para buscar a unificação. No entanto, Kitcher não parece endossar essa ideia e ela não se encaixa muito bem em sua ênfase na transmissão social de informações causais. Além disso, parece que mesmo a unificação inconsciente requer habilidades cognitivas muito sofisticadas (construção e comparação de diferentes sistematizações dedutivas, etc.), que são implausíveis de atribuir a muitos aprendizes causais, como crianças pequenas.

Uma interpretação natural das passagens citadas acima e outras em Kitcher (1989) é esta: um processo social de comparar sistematizações alternativas de crenças e extrair suas consequências dedutivas ocorre no nível da comunidade, com grupos de pessoas discutindo entre si sobre as quais as sistematizações dedutivas gerais unificam melhor as crenças da comunidade como um todo. Crenças causais particulares são justificadas no nível da comunidade por serem mostradas como parte da melhor sistematização geral das crenças da comunidade e, então, são transmitidas do estoque comum da comunidade aos indivíduos por meio de

um processo de transmissão social.

Um problema óbvio com esse cenário é que o processo de justificação em toda a comunidade ainda deve ser realizado de alguma forma por atores individuais. Se, como parece ser o caso, existem muitas sociedades que possuem uma quantidade substancial de conhecimento causal e explicativo, mas nas quais ninguém possui um conceito explícito ou claramente articulado de um argumento dedutivamente válido ou é muito hábil em extrair as consequências dedutivas de crenças ou possui versões explícitas dos conceitos de Kitcher de número e rigor dos padrões argumentativos, como exatamente as crenças da comunidade que refletem a operação dessas noções devem se formar? Se, como Kitcher admite, é psicologicamente irreal presumir que seres humanos individuais deliberada e autoconscientemente passem pelo processo de comparar sistematizações alternativas quando adquirem crenças causais por meio da experiência, por que é mais realista supor que esse processo ocorre de alguma forma por meio das interações de atores individuais no nível da comunidade[20] ?

Existe uma segunda dificuldade relacionada. Suponha, para fins de argumentação, que seja desejável ter um sistema de crença unificado no sentido de Kitcher, seja porque a unificação está conectada à explicação e esta é intrinsecamente valiosa ou porque a unificação está conectada a outros objetivos (por exemplo, confirmação) que são desejáveis. Ainda não é óbvio por que seria valioso ter um conjunto de crenças que são um subconjunto pequeno e adequado das crenças que compõem esse sistema unificado, que é o que a maioria das pessoas parece ter, dadas as visões de Kitcher sobre a transmissão do conhecimento causal. Lembre-se da imagem básica de Kitcher: quando eu adquiro a crença de que, digamos, se o sal é enfeitiçado é causalmente irrelevante para se ele se dissolve e de que se ele é colocado na água é causalmente relevante, eu adquiro um fragmento da sistematização geral S da comunidade. Mas, adicionando um fragmento de *S* ou mesmo uma série de fragmentos de S ao meu depósito de crenças pode não resultar em eu ter um sistema de crenças que seja unificado ou que facilite quaisquer objetivos epistêmicos associados à unificação. Claro, se eu acabar adicionando tudo ou a maioria de **S** para o meu depósito de crenças, terei nesse ponto um conjunto de crenças que é unificado e que traz consigo todos os benefícios

---

[20] Um problema paralelo se aplica à sugestão de que toda explicação tem uma estrutura **DN**, a menos que alguma versão da estratégia de estrutura oculta possa funcionar.

da unificação. Mas, como Kitcher concorda, não é realista supor que a maioria das pessoas possua algo parecido com a sistematização completa **S** que unifica melhor todas as crenças em sua comunidade. Isso parece ser verdade, por exemplo, de nossa própria comunidade epistêmica, na qual o conhecimento, especialmente o conhecimento científico, está altamente disperso dentre um pequeno grupo de especialistas e em que nenhuma mente de uma pessoa (e ainda menos a mente de um membro típico) contém ou atua de acordo com a sistematização que melhor une as crenças de toda a comunidade. De forma mais geral, parece improvável que as diferentes partes $\mathbf{B}_i$ da sistematização da comunidade S que vários indivíduos i adquirem por meio da transmissão cultural será, em cada caso, sistematizações altamente unificadas. Em suma, é um grande problema da história da transmissão cultural que é difícil ver como a unificação poderia ser cognitivamente ou praticamente valiosa a menos que caracterizasse os sistemas de crenças dos indivíduos e não apenas da comunidade. No entanto, tomar o tipo de unificação que Kitcher associa ao conhecimento causal e explicativo para caracterizar os sistemas de crenças individuais parece psicologicamente irrealista à primeira vista. Isso não quer dizer que não haja nenhuma maneira de dar sentido à aquisição de conhecimento causal no quadro unificacionista, mas muito mais precisa ser dito sobre como isso funciona.

A exposição mais detalhada da posição de Kitcher pode ser encontrada em Kitcher (1989). Salmon (1989, p. 94) contém uma discussão crítica da versão de Friedman da explicação unificacionista da explicação, mas termina defendendo uma "reaproximação" entre as abordagens unificacionistas e o próprio modelo mecânico causal de Salmon. Woodward (2003) contém críticas adicionais à versão de unificacionismo de Kitcher.

6. Teorias Pragmáticas da Explicação

6.1 Introdução

Apesar de suas muitas diferenças, os relatos de Hempel (concentrando-se agora apenas no modelo ND em vez do modelo IE), Salmon, Kitcher e outros discutidos acima compartilham amplamente uma concepção geral comum sobre o que o projeto de construção de uma teoria da explicação deve envolver e (em uma extensão considerável) que critérios tal teoria deve satisfazer para ter sucesso.

Vamos dizer que uma teoria da explicação contém elementos "pragmáticos" se (i) esses elementos exigirem referência irredutível a fatos sobre os interesses, crenças ou outras características da psicologia daqueles que fornecem ou recebem a explicação e/ou (ii) referência irredutível ao "contexto" em que ocorre a explicação (para saber o que isso significa, veja abaixo). Embora os escritores discutidos acima concordem que os elementos pragmáticos desempenhem algum papel na atividade de dar e receber explicações, eles presumem que há um núcleo não pragmático na noção de explicação que é a tarefa central de uma teoria da explicação capturar. Ou seja, presume-se que essa noção central pode ser especificada de uma forma que não requer referência a características da psicologia dos explicadores ou de seus públicos e pode ser caracterizada em termos de características que são não contextuais no sentido de que elas são suficientemente gerais, abstratas e "estruturais", que podemos vê-las como válidas em uma gama de explicações com diferentes conteúdos e em uma gama de contextos diferentes. Frequentemente, mas nem sempre, afirma-se que muitos aspectos dessas características podem ser capturadas formalmente, por meio de relações como implicação dedutiva ou relevância estatística. Além disso, esses escritores veem o objetivo de uma teoria da explicação em capturar a noção *correta* de explicação, como em "a (ou uma) explicação do efeito fotoelétrico é tal e tal" em oposição à noção de explicação sendo considerada explicativa por uma audiência particular ou não, uma questão que presumivelmente depende de considerações como a possibilidade de que a audiência compreende os termos em que a explicação é formulada. Finalmente, conforme observado na introdução deste artigo, os escritores nesta tradição *não* tinham como objetivo capturar todas as várias maneiras em que a palavra "explicação" é usada no português comum. Em vez disso, eles se concentraram em uma classe muito mais restrita de exemplos em que o que é de interesse é (algo como) explicar "por que" algum resultado ou fenômeno geral ocorreu, em oposição a explicar, por exemplo, o significado de uma palavra ou como resolver uma equação diferencial. A motivação para essa restrição é simplesmente o juízo de que uma teoria interessante e não trivial tem maior probabilidade de surgir se for restrita em escopo dessa maneira. Para facilitar a alusão, chamemos isso de concepção "tradicional" da tarefa de uma teoria da explicação.

Algumas ou todas dessas suposições e metas são rejeitadas em relatos pragmáticos da explicação, ou, como às vezes são chamados, "contextuais". Os primeiros colaboradores para esta abordagem incluem Michael Scriven (1962) e Sylvan Bromberger (1966), com declarações mais sistemáticas na década de 1980

devido ao aparecimento de van Fraassen (1980) e Achinstein (1983). Visto que nem sempre está claro quais são os pontos de desacordo entre as explicações pragmáticas e tradicionais, algumas observações orientadoras sobre isso serão úteis antes de nos voltarmos para os detalhes. Os defensores de abordagens pragmáticas para explicação normalmente enfatizam o ponto de que a possibilidade de que o fornecimento de um determinado corpo de informações a algum público produz compreensão ou um senso de inteligibilidade ou é apropriado ou esclarecedor para esse público depende do conhecimento prévio e dos interesses dos membros do público e de outros fatores que têm a ver com o contexto local. Por exemplo, uma explicação da deflexão da luz das estrelas pelo sol que apela às equações de campo da Relatividade Geral pode ser altamente esclarecedora para um físico treinado, mas ininteligível para um leigo por causa de sua formação. Fatores desse tipo são agrupados como "pragmáticos" e sua influência é considerada para ilustrar pelo menos uma maneira pela qual as considerações pragmáticas entram na noção de explicação.

Tomada em si, a observação que acabamos de descrever parece completamente incontroversa e não entra em conflito com abordagens de explicação que geralmente são vistas como paradigmaticamente tradicionais. De fato, como observado acima, escritores como Hempel e Salmon concordam explicitamente que a explicação tem uma dimensão pragmática no sentido que acabamos de descrever, na verdade, Hempel invoca o papel de fatores pragmáticos em vários pontos para abordar contraexemplos *prima-facie* ao modelo ND[21] . Isso sugere que, muitas vezes pelo

---

[21] Lembre-se em particular de que Hempel responde a supostos contraexemplos, que tentam mostrar que o modelo *ND* falha em capturar as características direcionais da explicação, alegando que essas características devem ser entendidas em termos de "pragmática", ao invés de em termos de características que uma teoria tradicional e não pragmática pode capturar. Essa resposta tem geralmente impressionado críticos como *ad hoc*. Um tanto ironicamente, como discutido abaixo, o diagnóstico de van Fraasen da fonte das assimetrias explicativas é o mesmo que o de Hempel, apesar do fato de que seu relato é fundamentalmente oposto ao de Hempel em muitos aspectos. Uma maneira de entender isso é que van Fraassen simplesmente generaliza a abordagem de Hempel das assimetrias explicativas, relegando todas as explicações ao reino da pragmática.

menos[22] , o que é peculiar sobre as abordagens pragmáticas da explicação não é apenas a mera ideia de que a explicação tem uma "dimensão pragmática", mas sim a afirmação adicional e muito mais forte de que o projeto tradicional de construção de um modelo de explicação perseguido por Hempel e outros até agora não teve sucesso e, talvez, esteja **fadado** ao fracasso, e que isso ocorre **porque** fatores pragmáticos ou contextuais desempenham um papel central e ineliminável na explicação de uma forma que resiste à incorporação em modelos do tipo tradicional. Nesta visão, muito do que é peculiar sobre relatos pragmáticos (incluindo os relatos de van Fraassen e Achinstein discutidos abaixo) é sua oposição aos relatos tradicionais e seu **diagnóstico** de porque as teorias falham - elas falham porque omitem elementos pragmáticos ou contextuais. Será importante manter esse ponto em mente no que se segue, porque existe uma certa tendência entre os defensores das teorias pragmáticas de argumentar como se a superioridade de sua abordagem fosse estabelecida simplesmente pela observação de que a explicação tem uma dimensão pragmática; em vez disso, parece mais apropriado pensar que a verdadeira questão é se as abordagens tradicionais são inadequadas em princípio por causa de sua negligência da dimensão pragmática da explicação.

Uma segunda questão diz respeito a uma ambiguidade importante na noção de “pragmático”. Em uma compreensão natural dessa noção, uma consideração

---

[22] Escrevi "muitas vezes" aqui devido à seguinte complicação: defensores de abordagens pragmáticas às vezes escrevem como se estivessem envolvidos em um projeto diferente do tradicional, no sentido de que seu objetivo é caracterizar uma noção de explicação segundo a qual, por exemplo, o mesmo corpo de informações pode ser explicativo para um público e não para outro, e o objetivo é especificar o que todos ou a maioria dos usos da palavra "explicação" têm em comum - ou seja, uma noção de explicação que não é o alvo de contas tradicionais. Pensando em relatos pragmáticos dessa maneira, pode-se considerá-los complementares e de forma alguma em conflito com os projetos tradicionais. Assim, não há razão, em princípio, para os defensores dos relatos tradicionais se oporem a tais projetos, embora os tradicionalistas possam ser céticos quanto a ser provável eles produzirem algo interessante. No entanto, como os defensores de relatos pragmáticos geralmente desejavam usar esses relatos para criticar relatos tradicionais, isso sugere que eles não concebem o que estão fazendo apenas da maneira conciliatória que acabamos de descrever. Este é um dos vários pontos em que uma maior clareza sobre os objetivos por parte dos tradicionalistas e de seus críticos pragmáticos seria salutar.

pragmática é aquela que tem a ver com utilidade ou usabilidade a serviço de algum objetivo ligado aos interesses humanos, onde esses interesses são em algum sentido relevante "práticos". Chame essa noção de "pragmática$_1$". Nesta interpretação, o modelo ND de Hempel pode ser corretamente caracterizado como uma teoria pragmática$_1$, ou como contendo elementos pragmáticos$_1$, uma vez que vincula intimamente as informações explicativas ao fornecimento de informações úteis para fins de predição, e predição certamente se qualifica como um objetivo pragmático. Por razões semelhantes, a teoria da explicação de Woodward (2003) também pode ser considerada uma teoria pragmática$_1$, uma vez que conecta a explicação com o fornecimento de informações que são úteis para manipulação e controle, objetivos inquestionavelmente úteis. Como esses exemplos sugerem, os modelos de explicação que aspiram a objetivos tradicionais podem ser teorias pragmáticas$_1$.

No contexto das teorias da explicação, entretanto, o rótulo "pragmático" geralmente tem a intenção de sugerir um conjunto um tanto diferente de associações. Em particular, "pragmático" é normalmente usado para caracterizar considerações que têm a ver com fatos sobre a psicologia (interesses, crenças, etc.) daqueles envolvidos em fornecer ou receber explicações e/ou para caracterizar considerações envolvendo o contexto local, muitas vezes com a sugestão que ambos os conjuntos de considerações podem variar de maneiras complexas e idiossincráticas que resistem à incorporação no tipo de teoria geral buscada pelos modelos tradicionais[23] . Chame este conjunto de associações de "pragmático$_2$". Nem a teoria de Hempel nem a de Woodward são pragmáticas$_2$. Em particular, como o exemplo do modelo ND ilustra, o fato de que uma teoria é pragmática$_1$ no sentido de que apela a fatos sobre objetivos geralmente compartilhados por seres humanos (como a previsão)

---

[23] Não estou ciente de qualquer exploração histórica sistemática das origens desse uso particular de "pragmático" em conexão com relatos de explicação. No entanto, uma conjectura (e é apenas uma conjectura) é que deriva, pelo menos em parte significativa, da linguística e, em particular, do contraste entre, por um lado, sintaxe e semântica, (muitas vezes pensadas como "objetivas" e objetos apropriados de estudo sistemático geral) e "pragmática" entendida como tendo a ver com o uso da linguagem por falantes específicos direcionados a públicos específicos em contextos específicos para atingir fins específicos. Esta noção de "pragmático" sugere uma conexão com o que é idiossincrático para a psicologia de usuários e contextos de linguagem específicos e contrasta com projetos que são considerados como fornecendo uma "sintaxe" ou "semântica" de explicação.

para ajudar a motivar um modelo de explicação **não** impede a tentativa de construir modelos de explicação que satisfaçam os objetivos tradicionais e não exige compromisso com a ideia de que a explicação deve ser entendida como uma noção pragmática$_2$. Precisamos ter o cuidado de distinguir essas duas maneiras diferentes de pensar sobre a dimensão "pragmática" da explicação.

Finalmente, como enfatizado acima, uma preocupação com a pragmática da explicação, naturalmente se conecta com um interesse na "psicologia" da explicação, e isso por sua vez sugere a relevância de estudos empíricos sobre tipos de informação que vários sujeitos (pessoas comuns, cientistas) achem explicativos, achem como fornecendo "entendimento", as distinções que os sujeitos fazem entre as explicações e assim por diante. Embora haja uma literatura crescente nessa área, os defensores filosóficos mais proeminentes das abordagens pragmáticas da explicação até agora tendem a não fazer uso dela. Nesse sentido, é importante ressaltar que essa literatura psicológica vai muito além dos truísmos encontrados na discussão filosófica sobre diferentes pessoas encontrarem diferentes tipos de informações explicativas, dependendo de seus interesses. Em particular, psicólogos têm se mostrado muito interessados em explorar características gerais ou padrões estruturais presentes em informações que vários sujeitos consideram explicativos. Por exemplo, Lombrozo (2010) encontra evidências de que sujeitos preferem explicações que apelam para relações que são relativamente estáveis, no sentido de continuar a se manter em circunstâncias variáveis[24] , e Lien e Cheng (2000) apresentam evidências de que, nos casos em que o *explanandum* **E** tem uma única causa candidata **C**, os sujeitos preferem níveis de explicação/descrição causal que maximizam **Δp** = **Pr**(**E**|**C**)−**Pr**(**E** | não-**C**).

Observe que, em ambos os casos, essas são relações ou padrões do tipo que as explicações tradicionais tentam capturar. Como esses exemplos mostram, não há incompatibilidade necessária entre o projeto de tentar formular uma explicação que satisfaça os objetivos tradicionais e o interesse pela psicologia da explicação. Pode ser o caso de que os sujeitos considerem certos tipos de informação explicativos ou produtores de compreensão **porque** certas características estruturais do tipo que os relatos tradicionais tentam caracterizar estão presentes nessas informações, na verdade, é isso que os artigos de Lombrozo e Lien e Cheng sugerem.

---

[24] A noção relevante de estabilidade é a noção discutida em Woodward (2006).

Na mesma linha, também devemos distinguir o projeto geral de investigação da psicologia empírica da explicação (que pode ser buscado por meio de uma variedade de compromissos diferentes sobre como melhor teorizar sobre a explicação) da afirmação mais específica de que a caracterização de para que serve uma relação explicativa entre *explanans* e *explanandum* deve ser dada em termos "psicologísticos" no sentido de que isso requer referência irredutível a fatos psicológicos sobre públicos específicos, como os caprichos daquilo em que eles estão interessados. Em geral, a possibilidade de haver regularidades robustas conectando características estruturais ou objetivas em corpos de informação com a possibilidade de essa informação ser julgada como explicativa por vários assuntos deve ser considerada uma questão empírica e não como algo que pode ser resolvido da poltrona. **Pode** ser o caso de que não existam tais regularidades e que o que as pessoas consideram explicativo ou gerador de entendimento varia enormemente, dependendo de seus interesses e de outros fatores psicológicos, mas isso é algo que precisa ser mostrado, não presumido no início da investigação.

6.2 Empirismo Construtivo e Teoria Pragmática da Explicação

Um dos relatos pragmáticos de explicação mais influentes está associado ao empirismo construtivo. Esta é a tese defendida por Bas van Fraassen em seu livro, de 1980, *A Imagem Científica*, em que o objetivo da ciência (ou pelo menos ciência "pura") é a construção de teorias que sejam "empiricamente adequadas" (isto é, que produzam uma descrição verdadeira ou correta de observáveis) e não, como os realistas científicos supõem, teorias que pretendem contar histórias literalmente verdadeiras sobre inobserváveis. Da mesma forma, a “aceitação” de uma teoria envolve apenas a crença de que ela é empiricamente adequada (VAN FRAASSEN, 1980, p. 12). O relato da explicação de van Fraassen, que é apresentado em vários artigos e, mais completamente, no capítulo seis de seu (1980), pretende se ajustar a essa concepção geral da ciência: é uma concepção segundo a qual a explicação *per se* não é um objetivo epistêmico da ciência “pura” (a adequação empírica é o único objetivo), mas sim uma virtude “pragmática”, que tem a ver com a “aplicação” da ciência (observe que, na medida em que a aplicação da ciência é considerada uma questão "pragmática" (ou seja, pragmática$_1$) e a ideia de que a explicação é pragmática nesse respeito é usada para motivar a adoção de uma teoria da explicação pragmática (isto é, pragmática$_2$), temos uma transição entre as duas noções de

"pragmático" distinguidas acima). Como explicação é uma virtude meramente pragmática, uma preocupação com a explicação não é algo que pode exigir dos cientistas ir além da crença na adequação empírica de suas teorias para a crença na verdade literal das afirmações sobre entidades inobserváveis.

De acordo com van Fraassen, explicações são respostas a perguntas e obter clareza sobre a lógica das perguntas é central para construir uma teoria da explicação. As perguntas podem assumir muitas formas diferentes, mas quando a pergunta de interesse é uma pergunta do tipo "por que", as perguntas explicativas geralmente assumem a seguinte forma: uma pergunta sobre o motivo de algum *explanandum* $\mathbf{P_k}$ em vez de qualquer um dos membros de um contraste **X** (um conjunto de alternativas possíveis para $\mathbf{P_k}$) obtido. Além disso, alguma "relação de relevância" **R** é assumida pela pergunta. Uma **resposta A** a esta pergunta assumirá a forma "$\mathbf{P_k}$ em contraste com (o resto de) **X** porque **A**, onde **A** carrega a relação de relevância **R** para [$\mathbf{P_k}$, **X**]". Para usar o exemplo de van Fraassen, considere "Por que este condutor está deformado?" Dependendo do contexto, o contraste pretendido pode ter a ver com, por exemplo, porque este condutor em particular está deformado em contraste com algum outro condutor que não está deformado ou, alternativamente, pode ter a ver com porque este condutor específico está deformado agora quando estava anteriormente não distorcido. A relação de relevância **R** da mesma forma depende do contexto e das informações que o questionador está interessado em obter. Por exemplo, **R** pode envolver informações causais (a questão pode ser uma solicitação sobre o que causou a deformação), mas também pode ter a ver com informações sobre a função, se o contexto for aquele em que se presume que a forma do condutor desempenha algum papel funcional em uma estação de energia sobre a qual o questionador deseja saber. Assim, "contexto" entra na explicação desempenhando um papel na especificação da classe de contraste **X** e a relação de relevância **R**. van Fraassen descreve várias regras para a "avaliação" das respostas. Por exemplo, $\mathbf{P_k}$ e **A** devem ser verdadeiros, os outros membros da classe de contraste não devem ser verdadeiros, **A** deve "favorecer" (aumentar a probabilidade condicional de) $\mathbf{P_k}$ contra alternativas, e **A** deve comparar-se favoravelmente com outras respostas à mesma pergunta, uma condição que tem vários aspectos, incluindo, por exemplo, se *A* favorece o tópico mais do que essas outras respostas e se **A** é filtrado por outras respostas. No entanto, ele também deixa claro (como o exemplo acima sugere) que uma variedade de relações de relevância diferentes pode ser apropriada dependendo do contexto e que a avaliação das respostas também depende do contexto. Além disso, ele nega explicitamente

que haja algo distintivo na categoria de explicação científica que tem a ver com sua estrutura ou forma, em vez disso, uma explicação científica é simplesmente uma explicação que faz uso de informações que são (ou pelo menos que são tratadas como) com base em uma teoria "científica".

Van Fraassen resume sua visão da explicação (e aponta seus fundamentos para rejeitar abordagens objetivistas) como a seguir:

> A discussão da explicação deu errado já no início, quando explicação foi concebida como uma relação semelhante à descrição: uma relação entre uma teoria e um fato. Na verdade, é uma relação de três termos entre teoria, fato e contexto. Não é de admirar que nenhuma relação única entre teoria e fato tenha conseguido caber mais do que em alguns exemplos! Ser uma explicação é essencialmente relativo, porque uma explicação é uma resposta [...] ela é avaliada frente a uma pergunta, que é um pedido de informação. Mas exatamente [...] o que é pedido difere de contexto para contexto (1980, p. 156).

Van Fraassen começa seu capítulo sobre explicação com uma breve história que fornece um bom ponto de entrada para como ele pretende que seu relato funcione. Lembre-se da seção 2.5 que um contraexemplo bem conhecido para o modelo ND envolve a alegação de que se pode explicar o comprimento da sombra projetada S por um mastro em termos de altura **H** do mastro, mas que, supostamente, não se pode explicar **H** em termos de **S**, apesar do fato de que se pode construir uma derivação de ND a partir de **S** para **H**. Isso é comumente considerado para mostrar que o modelo ND deixou de fora algum fator relacionado às características direcionais ou assimétricas da explicação - por exemplo, talvez uma assimetria na relação entre causa e efeito que deveria ser incorporada ao modelo de explicação de alguém. Na história de van Fraassen, uma explicação causal direta do tipo usual de **S** em termos de **H** (embora o objeto em questão seja uma torre ao invés de um mastro) é oferecida primeiro. Em seguida, uma segunda explicação, segundo a qual a altura da torre é "explicada" pelo fato de ter sido projetada para lançar uma sombra de determinado comprimento, é avançada. Presumivelmente, a moral que devemos tirar é aquela que, conforme o contexto e talvez a relação de relevância **R** são variados, ambos

H explica **S**
e
S explica **H**

são explicações aceitáveis (legítimas, apropriadas, etc.). Além disso, uma vez que essas variações no contexto e na relação de relevância geram variações no que é de interesse do explicador e de seu público, somos ainda encorajados a concluir que as assimetrias explicativas têm sua origem em fatos psicológicos sobre os interesses das pessoas e crenças de fundo, ao invés de, digamos, alguma assimetria que existe na natureza independentemente destes. Pragmatistas sobre a explicação pensam que uma conclusão semelhante se aplica a outras características da relação de relevância explicativa que os filósofos tentaram caracterizar em termos de modelos tradicionais de explicação.

Uma resposta óbvia a esta afirmação, feita por vários críticos (KITCHER; SALMON, 1987, p. 317), é que o exemplo não envolve realmente um caso em que, dependendo do contexto, **H** explica causalmente **S** e **S** explica causalmente **H**. Em vez disso, embora H explica causalmente em **S**, é (algo como) o **desejo** por uma sombra de comprimento **S** (ao invés de **S** em si) que explica (ou pelo menos explica causalmente) a altura (ou a escolha da altura) para a torre. Ou, se preferirmos, no último caso, recebemos algo como uma explicação funcional (mas não uma explicação causal) para a altura da torre, no sentido de que somos informados qual é a função pretendida dessa escolha de altura. Em qualquer um desses diagnósticos, este não será um caso em que **H** fornece uma explicação causal de **S** ou se **S** ao invés fornece uma explicação causal de **H** que muda dependendo de fatores relacionados aos interesses do locutor ou público ou outros fatores contextuais. Se esse é o caso, a história sobre a torre não mostra que a assimetria presente no exemplo do mastro deve ser considerada em termos de fatores pragmáticos. Ela pode ser considerada de alguma outra maneira. Na verdade, embora a discussão deva estar além do escopo deste ensaio, uma série de possíveis candidatos para tal explicação não pragmática das assimetrias causais foram propostas, tanto na filosofia quanto fora dela (por exemplo, na literatura de aprendizado de máquina). Esses candidatos incluem assimetrias na conectabilidade causal do tipo descrito em Hausman (1998), assimetrias estatísticas de vários tipos (SPIRTES; GLYMOUR; SCHEINES, 2000) e assimetrias na dependência informacional (JANZING, 2012). Todas essas propostas podem estar erradas, mas é difícil ver como se mostra que estão erradas apenas pelos tipos de observações avançadas por van Fraassen na história da torre e das

sombras, mas, ao contrário, mostrar que estão erradas exigiria críticas detalhadas das próprias propostas[25] .

Uma crítica muito mais geral foi feita contra a versão de van Fraassen de uma teoria pragmática por Salmon e Kitcher (1987). Basicamente, a reclamação deles é que a relação de relevância **R** no relato de van Fraassen é completamente irrestrita, com a consequência (o que eles consideram obviamente inaceitável) de que para qualquer par de proposições verdadeiras ***P*** e ***A***, a resposta ***A*** é relevante para ***P*** por meio de alguma relação de relevância e, portanto, "explica" ***P***. Por exemplo, de acordo com Salmon e Kitcher, podemos definir uma relação de "influência astral" ***R****, atendendo aos critérios de van Fraassen para ser uma relação de relevância, de modo que o tempo ***t*** da morte de uma pessoa é explicada em termos de ***R**** e a posição de vários corpos celestes em ***t***. Aqui pode parecer que van Fraassen tenha uma resposta pronta. Como observado acima, na visão de van Fraassen, o conhecimento prévio e, no caso da explicação científica, o conhecimento científico atual, ajuda a determinar quais são as relações de relevância aceitáveis e as respostas aceitáveis para as questões colocadas nos pedidos de explicação - tais conhecimentos e expectativas que a acompanham fazem parte do contexto relevante quando se pede uma explicação sobre a hora da morte. Obviamente, a influência astral não é uma relação de relevância aceitável ou legítima de acordo com a ciência moderna - portanto, o apelo a tal relação não é considerado explicativo pela teoria de van Fraassen. De forma mais geral, pode-se argumentar que o conhecimento científico disponível fornecerá restrições nas relações de relevância e respostas que excluem a preocupação "vale tudo" levantada por Salmon e Kitcher - pelo menos na medida em que o contexto é aquele em que uma "explicação científica" é buscada.

Embora esta resposta possa parecer plausível até certo ponto, ela mostra até que ponto muito do trabalho de distinguir o explicativo do não explicativo no relato de van Fraassen vem de um apelo muito geral ao que é aceito como informação prévia legítima na ciência atual. Em outras palavras, isso levanta a preocupação

---

[25] Este exemplo também nos lembra que as tentativas de fornecer caracterizações tradicionais de características de relações explicativas (ou pelo menos causais) não se limitam de forma alguma a filósofos - tentativas de fazer isso também podem ser encontradas no aprendizado de máquina, como em Janzing et al. (2012), e em teorias normativas de juízo causal propostas por psicólogos, como em Lombrozo (2010) e Cheng (2000).

de que, uma vez que se mova além da maquinaria formal de van Fraassen sobre perguntas e respostas (que o próprio van Fraassen reconhece ser relativamente irrestrito), fica-se com um relato segundo o qual uma explicação científica é simplesmente qualquer explicação empregando afirmações de ciência atual e uma relação de relevância atualmente e cientificamente aprovada. Mesmo que de outra forma irrepreensível, esta proposta é, se não exatamente trivial, pelo menos bastante deflacionária - ela fornece muito menos do que muitos esperavam de uma teoria da explicação. Em particular, em casos (dos quais há muitos exemplos) em que há uma discussão ou disputa em andamento em alguma área da ciência não sobre se alguma teoria ou modelo proposto é verdadeiro, mas sim sobre se explica algum fenômeno, não é fácil ver como a proposta ainda pretende fornecer orientação. Por outro lado, a réplica óbvia que pode ser feita, em nome de van Fraassen, é que nenhuma abordagem mais ambiciosa, que satisfaça as expectativas associadas a explicações mais tradicionais, é possível (incluindo uma demarcação de explicações candidatas em aquelas que são "corretas" e "incorretas"), uma teoria como a de van Fraassen é a melhor possível. Se não há uma teoria defensável da explicação que incorpore uma relação de relevância não trivialmente restritiva, não pode ser uma boa crítica à teoria de van Fraassen que ele falhe em fornecer isso[26] .

Portanto, pelo menos da perspectiva de van Fraassen, os modelos tradicionais não estão em melhor posição do que o seu para fornecer tal orientação.

Um último ponto sugerido pela teoria de van Fraassen é este. Ao considerar as teorias pragmáticas, importa muito onde exatamente os elementos “pragmáticos” são reivindicados para entrar no relato da explicação. Um ponto em que tais considerações parecem entrar claramente é na seleção ou caracterização do que o público deseja que seja explicado. Isso se reflete na teoria de van Fraassen na escolha de um $\mathbf{P}_k$ e uma classe de contraste associada **X**. Obviamente, se estamos procurando uma explicação de porque, digamos, este condutor em particular agora está dobrado quando antes estava reto ou se, em vez disso, queremos saber por que esse condutor está dobrado enquanto outro condutor está reto, isso é uma

---

[26] Em sua discussão muito lúcida das teorias pragmáticas da explicação, Salmon (1989) enfatiza o fracasso tanto das teorias de van Fraassen quanto de Achinstein em fornecer uma caracterização de uma relação de relevância explicativa objetiva. Embora Salmon esteja correto sobre isso, suas observações não se ocupam totalmente das visões de van Fraassen e Achinstein de que não existe tal caracterização.

questão que depende de nossos interesses. No entanto, este tipo particular de "relatividade de interesse" (e fenômenos associados que têm a ver com o papel do foco contrastivo na caracterização de explananda, que na verdade apenas servem para especificar mais exatamente quais explananda particulares queremos explicados) parece algo que pode ser prontamente reconhecido pelas teorias tradicionais[27]. Afinal, não é uma ameaça ao ND ou outros modelos com aspirações tradicionais semelhantes de que um público pode estar interessado em uma explicação do efeito fotoelétrico, mas não a deflexão da luz das estrelas pelo sol e outro público pode ter interesses opostos. O que seria uma ameaça para o modelo ND e modelos semelhantes seria um argumento que uma vez que algum *explanandum* **E** é totalmente especificado, se o *explanans* **M** explica **E** (isto é, se existe uma relação explicativa entre **M** e **E**) é ela própria "relativa a interesses". É natural interpretar van Fraassen fazendo esta última afirmação, tanto em conexão com as assimetrias explicativas quanto de forma mais geral.

6.3 Explicando como um Ato Ilocucionário

Outra abordagem pragmática de explicação muito influente se concentra no ato de explicar e trata isso como um ato Ilocucionário, no sentido em que essa

---

[27] Um ponto intimamente relacionado é que uma caracterização que "relativiza" alguma característica de uma explicação para um contexto às vezes pode, por assim dizer, ser "desrelativizada" tornando explícito como a característica em questão depende do contexto - em outras palavras, a aparente contextualidade pode ser apenas um reflexo do fato de que alguma característica relevante não foi explicitada. Por exemplo, um explanandum para o qual a classe de contraste é deixada implícita pode parecer apoiar uma imagem de explicação relativa ao contexto (por exemplo, que diferentes explicações de porque o condutor está dobrado serão apropriadas dependendo do contexto), mas pelo menos às vezes esta aparência pode ser removida tornando explícita a classe de contraste pretendida. Pode-se argumentar que, uma vez que fique explícito que o que interessa é porque este condutor está dobrado enquanto os outros estão retos, é uma questão "objetiva" se algum candidato a explanans explica esse contraste. Por esta razão, parece que devemos considerar uma teoria pragmática completa como aquela que (como a de Achinstein e presumivelmente de van Fraassen) afirma que as explicações têm um elemento contextual que não pode ser removido (de uma forma que satisfaça as restrições objetivistas) ao se tornar o contexto explícito.

noção é usada na teoria dos atos de fala. A exposição mais sistemática dessa abordagem deve-se a Peter Achinstein (ver especialmente, Achinstein, 1983). Como muitos outros teóricos pragmáticos, Achinstein está interessado em capturar uma noção muito ampla de explicação, que inclui não apenas explicações causais (e não apenas respostas a perguntas questionadoras), mas noções como explicar o significado da palavra, as regras do xadrez, a função de alguma estrutura biológica e assim por diante. Seu relato é complexo demais para ser descrito em detalhes; tudo o que pode ser tentado aqui é um esboço muito bruto.

O ponto de partida de Achinstein é o que está envolvido em alguém explicar algo a outro. De acordo com Achinstein, em tais episódios, a intenção da pessoa que explica é crucial: em particular, ao explicar, o explicador deve ter a intenção de tornar algo (aproximadamente, um certo tipo de pergunta indireta **q** correspondendo a algo explicado) "compreensível". Uma explicação (entendida como o produto de um ato explicativo) é então definida como um par ordenado, "em que um de seus membros é um tipo de ato explicativo **q** (como acima) [...] e cujo outro membro é uma proposição que fornece uma resposta à pergunta **q**." (2010, p. XI). Por exemplo, a explicação de Newton de porque as marés ocorrem é representada pelo par ordenado: as marés ocorrem por causa da atração gravitacional da lua: explicando por que as marés ocorrem. Achinstein distingue entre explicações "corretas" e "boas". "Uma explicação correta é aquela em que o membro proposicional do par ordenado é verdadeiro" (2010, p. XI). Uma explicação correta pode, não obstante, não ser boa porque, por exemplo, é inadequada de várias maneiras para as habilidades e interesses do público ao qual é dirigida. A noção de uma boa explicação é ainda caracterizada em termos de um conjunto de instruções para a construção da explicação, onde essas instruções são sensíveis aos interesses, crenças e assim por diante do público. Essas instruções podem especificar, por exemplo, que se busca uma explicação causal, em vez de algum outro tipo de explicação, ou que a explicação buscada deve fazer referência a microentidades. Uma característica muito importante da posição "pragmática" de Achinstein é que não existe um único conjunto universal de instruções que seja apropriado para todos os públicos e contextos, seja na ciência ou em outro lugar. Assim, os relatos tradicionais que pretendem fornecer tais instruções estão (neste aspecto) errados. Achinstein escreve

> Agora, deixe-me fazer uma conjectura. Suponha que, seguindo os passos de Hempel e Salmon, você formule um conjunto de critérios objetivos e não

> pragmáticos que você acha que todas as explicações científicas devem satisfazer para serem bem avaliadas. Esses critérios serão universais no sentido de que não devem variar de uma explicação para a outra, mas devem ser aplicáveis a todas as explicações científicas. Eles também são universais no sentido de que não devem incorporar suposições ou pressuposições empíricas específicas que possam ser feitas por cientistas em um campo ou contexto, mas não em outro. Portanto, eles podem incluir o uso de leis, fatores causais e hipóteses quantitativas, a satisfação de algum critério de unificação ou simplicidade, e assim por diante. Minha conjectura é que qualquer conjunto de objetivos não pragmáticos, critérios universais que você proponha, você será capaz de encontrar ou construir contraexemplos a ele, mas como um conjunto de condições necessárias e como um conjunto de condições suficientes (2010, p. 137).

Achinstein ilustra essa afirmação com referência à explicação de Rutherford, de 1911, sobre o espalhamento de partículas alfa. A explicação de Rutherford apelou para suposições sobre a estrutura atômica, em particular, que a carga positiva de um átomo está concentrada em seu núcleo cujo volume é pequeno em comparação com o volume total do átomo, para derivar uma expressão quantitativa para a magnitude do espalhamento em vários ângulos. De acordo com Achinstein, outras explicações concorrentes (por exemplo, uma explicação que apenas deu a expressão quantitativa que governa a dispersão, mas não conectou isso a afirmações sobre a estrutura atômica) podem satisfazer os vários critérios tradicionais para qualidade explicativa encontrados na literatura de filosofia da ciência (tais explicações pode ter uma estrutura ND, descrever as causas, ser unificadoras, etc.), mas, mesmo assim, serão menos boas do que a de Rutherford. A explicação de Rutherford é boa (ou tão boa quanto é) porque fornece uma explicação "*no nível subatômico da matéria de uma forma que os físicos da época estavam interessados em explicar a dispersão*" (2010, p. 136, grifo do autor). Em outras palavras, para explicar os aspectos em que a explicação de Rutherford é boa (ou melhor do que as concorrentes), devemos fazer referência irredutível aos interesses dos físicos da época. Nesse

sentido, o relato de Achinstein de “boa explicação” é, como ele diz, “fortemente pragmático”. Sua “conjectura” captura muito bem (embora talvez não tudo) o que está em questão entre relatos de explicação pragmáticos e tradicionais, não puramente não-pragmáticos. A questão central é se podemos capturar os aspectos em que a explicação de Rutherford é melhor do que alternativas em termos de uma relação explicativa que pode ser especificada independentemente dos interesses (e talvez de outros fatos psicológicos relativos) a públicos específicos e também independentemente de fatos irredutivelmente "contextuais” (como a alegação de que, neste caso, uma boa explicação requer referência ao nível subatômico, mas não há nada mais geral a ser dito, independentemente dos fatos sobre os interesses das pessoas, sobre por que as explicações neste nível são preferíveis). Um argumento convincente para a segunda alternativa presumivelmente precisaria (pelo menos) examinar os relatos tradicionais existentes e mostrar que eles não tiveram sucesso. Esse é um projeto que Achinstein empreende em seu (1983), sem surpresa, os juízos sobre se ele consegue defender o fracasso das explicações objetivistas diferem.

6.4 Reflexões finais sobre teorias pragmáticas

Até agora, tratamos os relatos “pragmáticos” e “tradicionais” como possibilidades diametralmente opostas. Isso corresponde à forma como esses relatos são normalmente apresentados, tanto por seus defensores quanto por seus detratores. No entanto, vale a pena (e fornece uma visão adicional sobre ambas as abordagens), considerar a possibilidade de que algumas das ideias associadas a cada uma delas possam ser combinadas, ampliando assim o espaço de possíveis abordagens para a explicação. Em primeiro lugar, podemos distinguir a afirmação de que a explicação tem elementos irredutivelmente “contextuais” da afirmação de que esses elementos contextuais devem ser entendidos em termos de fatos sobre a psicologia (interesses, etc.) das partes da explicação. Uma possibilidade alternativa é que a explicação é de fato irredutivelmente contextual, mas que esses elementos contextuais devem ser entendidos não psicologicamente - grosso modo, em termos do papel de fatos empíricos particulares na explicação, onde a relevância desses fatos resiste à captura por meio dos recursos empregados no modelo ND e outros modelos tradicionais (isto é, resiste à captura em termos de relações, como implicação dedutiva, relevância estatística e assim por diante). Uma possível ilustração é fornecida pelo próprio exemplo de Achinstein da explicação da dispersão de partículas

alfa em termos de fatos sobre estrutura nuclear. Como observado acima, o próprio Achinstein pensa que, neste caso, quão boa é a explicação é dependente do contexto, porque depende de fatos psicológicos: em particular, quão boa é a explicação reflete o fato de que os físicos estão particularmente "interessados" em explicações que apelam à estrutura nuclear. Uma possibilidade alternativa é que a explicação tenha elementos irredutivelmente "contextuais" no sentido de que há algo sobre os detalhes empíricos deste caso particular que torna os fatos sobre a estrutura nuclear explicitamente relevantes para a dispersão, mas onde essa relevância não pode ser totalmente capturada em termos de termos das características abstratas e estruturais (os critérios "objetivos", não pragmáticos e universais de Achinstein) nos quais os modelos tradicionais de explicação se concentram. Assim, essa possibilidade envolve uma noção não psicológica de "contextual" que contrasta com a ideia de que a relação explicativa pode ser especificada de uma forma "independente do conteúdo". Para esclarecer essa noção de independência de conteúdo, considere o modelo ND. Este modelo é independente de conteúdo, no sentido de que afirma que, enquanto uma certa relação estrutural abstrata se mantém entre os *explanans* (**C_i**, **L_i**) e o *explanandum* **E**, não importa o que especificamente se preenche **C_i**, **L_i** e **E**, a estrutura resultante é uma explicação. Um contextualista sobre a explicação no sentido não psicológico afirmaria, em vez disso, que para qualquer candidato independente de conteúdo para a relação explicativa que especificamos (seja essa especificação em termos de relações dedutivas ou probabilísticas ou qualquer outra coisa similarmente formal e abstrata ou com aspirações semelhantes à universalidade) haverá exemplos que instanciam essa estrutura que são explicativos e exemplos que não são explicativos - neste sentido, o conteúdo específico que preenchemos para o candidato explica a questão de saber se temos uma explicação.

Uma analogia adicional pode ajudar a concretizar essa ideia. John Norton (no prelo) defendeu em uma série de artigos o que chama de "teoria material da indução". Sua visão é que a confiabilidade de várias inferências indutivas depende de suposições empíricas específicas associadas ("materiais") de uma forma que impede a formulação de qualquer lógica universal de indução - não há forma universal de um argumento indutivo que garanta confiabilidade, independentemente do conteúdo específico do que entra nesse argumento. No entanto, não faz parte da visão de Norton que o suporte indutivo é de alguma forma uma questão subjetiva ou relativa aos interesses etc. de públicos específicos - são os fatos empíricos de tipo não psicológico (exceto, é claro, quando as relações de evidência de interesse dizem

respeito a hipóteses psicológicas) que sustentam as relações de evidência. Podemos dizer que, em sua visão, a inferência indutiva é “contextual” no sentido de que não é independente de conteúdo, mas também não requer uma caracterização psicologística.

A possibilidade em consideração é que uma afirmação semelhante pode ser verdadeira para explicação. Talvez seja verdade, por exemplo, que, a fim de capturar os aspectos em que a explicação de Rutherford é boa, seja necessário invocar, além do conteúdo relativamente geral, requisitos independentes sobre unificação, derivabilidade de leis e, assim por diante, restrições que tem a ver com fatos materiais "locais" mais específicos sobre a estrutura atômica, no sentido de que nada contará como uma explicação da dispersão de alfa que não invoque tais fatos, e que outras explicações com a mesma forma que não invoquem estrutura atômica não contarão como explicativas. Mas talvez essas restrições adicionais tenham a ver com fatos sobre como o mundo é, e não, como Achinstein sugere, com fatos sobre o que os físicos da época estavam mais interessados. Uma visão desse tipo pode capturar (ou conceder) algumas das afirmações feitas por abordagens pragmáticas sobre o papel dos elementos contextuais na explicação, mas evitaria algumas das tendências subjetivas ou psicologistas em tais abordagens. Seria “contextual” no sentido de que a teoria material da indução de Norton é contextual.

Um pensamento intimamente relacionado é que se alguém está inclinado a incorporar elementos contextuais na teoria da explicação, permanece uma gama de possibilidades sobre como eles podem ser combinados com elementos mais universalistas. Como sugerido acima, em vez de pensar nesses dois conjuntos de elementos como simplesmente opostos um ao outro, pode ser melhor pensar em termos dos dois trabalhando juntos de forma sinérgica. Como ilustração, considere a noção de unificação. Pode ser que não possamos fornecer uma caracterização adequada desta noção e seu papel na explicação em termos puramente formais, completamente independentes de conteúdo - por exemplo, em termos como derivar muitas conclusões de algumas suposições básicas ou substituir teorias com muitos parâmetros livres por teorias que têm apenas alguns desses parâmetros. Não obstante, pode ser verdade que, uma vez que o material local ou as restrições empíricas são usadas para restringir a classe de teorias candidatas a serem comparadas com relação à unificação que alcançam, algo como contar suposições básicas ou número de parâmetros livres (ou mais plausivelmente algo no mesmo espírito, mas mais sofisticado) fornece informações úteis sobre o grau de unificação alcançado. Novamente, a analogia com as teorias do raciocínio indutivo é sugestiva. Certamente não é o caso de que todas as tentativas de fornecer teorias formais ou gerais nesta área sejam equivocadas ou fadadas ao fracasso, as vários

abordagens de inferência estatística e aprendizado de máquina são contraexemplos óbvios a essa sugestão. Por outro lado, as teorias de sucesso nessa área não são completamente universais ou independentes de conteúdo; em vez disso, elas, em muitos casos, produzem resultados que parecem sensatos ou normativamente corretos em uma certa gama de aplicações ou quando certas condições empíricas de fundo são satisfeitas, mas não em outras situações. Em outras palavras, essas teorias são sensíveis ao contexto e contêm elementos que parecem objetivos e estruturais. Talvez algo assim acabe sendo verdadeiro para a "explicação".

Van Fraassen (1980), especialmente o capítulo seis e Achinstein (1983) são defesas clássicas de abordagens pragmáticas para a explicação. Essas teorias pragmáticas são discutidas e criticadas por Salmon (1989). A teoria de van Fraassen também é discutida em Kitcher e Salmon (1987). De Regt e Dieks (2005) trazem uma defesa recente do que os autores descrevem como uma tese "contextual" da compreensão científica e que envolve alguns dos temas da literatura da "pragmática da explicação".

7. Conclusões, questões em aberto e direções futuras

O que podemos concluir dessa exposição de algumas das tentativas recentes mais proeminentes de construir modelos de explicação científica? Que questões importantes permanecem em aberto e quais são as direções mais promissoras para trabalhos futuros? É claro que qualquer esforço de balanço refletirá um ponto de vista específico, mas com essa advertência em mente, várias observações parecem plausíveis, mesmo que não totalmente incontroversas.

7.1 O papel da causalidade

O primeiro diz respeito ao papel da informação causal na explicação científica. É um juízo plausível, embora de forma alguma inevitável[28] que muitas

[28] A aceitação dessa avaliação dependerá em parte de se pensar que existem formas não causais de explicação-porque (no sentido amplo e vago de explicação-porque apontada na Seção 1). Se houver, então uma teoria da causalidade mais adequada nos levará apenas parte do caminho em direção a uma teoria da explicação mais satisfatória.

das dificuldades enfrentadas pelos modelos descritos acima derivam de sua confiança no que parecem ser tratamentos inadequados de causalidade e relevância causal. Os problemas de assimetrias explicativas e irrelevância explicativa descritos na seção 2.5 parecem mostrar que a posse de uma lei (entendida como uma regularidade) entre **C** e **E** não é suficiente para **C** causar **E**; portanto, não é uma condição suficiente para **C** figurar em uma explicação de **E**. Se o argumento da seção 3.3 estiver correto, o problema fundamental com o modelo RE é que as informações de relevância estatística são insuficientes para capturar totalmente as informações causais, no sentido de que diferentes estruturas causais podem ser consistentes com as mesmas informações sobre relações de relevância estatística. Da mesma forma, o modelo MC enfrenta a dificuldade de que as informações sobre os processos e interações causais também são insuficientes para capturar totalmente as relações de relevância causal e que há uma variedade de casos em que as relações causais se mantêm entre **C** e **E** (e, portanto, em que **C** figuras em uma explicação de **E**) embora não haja nenhum processo causal de conexão entre **C** e **E**. Finalmente, um problema fundamental com os modelos unificacionistas é que o conteúdo de nossos juízos causais não parece resultar de nossos esforços de unificação, pelo menos quando a unificação é entendida ao longo das linhas defendidas por Kitcher. Por exemplo, como discutido acima, as considerações que têm a ver com a unificação não explicam por si mesmas por que é apropriado explicar os efeitos em termos de suas causas e não vice-versa.

No mínimo, essas observações sugerem que o progresso em conexão com a "explicação científica" pode exigir mais atenção à noção de causalidade e uma integração mais completa das discussões de explicação com a literatura emergente sobre causalidade, tanto dentro como fora da filosofia[29] . Relatos contrafactuais de causalidade podem ser promissores neste contexto (WOODWARD, 2003).

Isso significa que um foco na causalidade deve substituir inteiramente o projeto de desenvolver modelos de explicação ou que os filósofos devem parar de falar sobre explicação e, em vez disso, falar apenas sobre causalidade? Apesar da centralidade da causalidade na explicação, é discutível que incluir completamente a última na primeira perde conexões com algumas questões importantes. Por um lado, as próprias afirmações causais parecem variar muito na medida em que são

---

[29] Trabalhos recentes relevantes incluem Hall (no prelo), Lewis (1986, 2000), Pearl (2000), e Spirtes, Glymour e Scheines (1993, 2000).

explicativamente profundas ou esclarecedoras. Afirmações causais encontradas na mecânica newtoniana parecem mais profundas ou mais satisfatórias do ponto de vista da explicação do que afirmações causais da variedade “a pedra quebrou a janela”. Em geral, supõe-se que tais diferenças estão conectadas a outras características, por exemplo, até que ponto é geral, estável e coerente com o conhecimento de fundo uma afirmação causal. Contudo, como observamos, nem todos os tipos de generalidade, estabilidade etc. parecem explicitamente relevantes (ou conectados à bondade explicativa). Portanto, mesmo que se foque apenas na explicação causal, permanece o importante projeto de tentar compreender melhor que tipo de distinções entre as afirmações causais importam para o bem na explicação. Na medida em que isso ocorre, os tipos de preocupações que animam os tratamentos tradicionais de explicação não parecem ser inteiramente subsumíveis nas descrições padrão de causalidade.

Há também a importante questão de se todas as formas legítimas de explicação por que são causais. Por exemplo, alguns escritores (NERLICH, 1979) afirmam que existe uma variedade de explicação física que é "geométrica" ao invés de causal, no sentido de que consiste em explicar fenômenos apelando para a estrutura do espaço-tempo ao invés de fatos sobre forças ou transferência de energia/momento (Nerlich considera que as explicações causais em física têm a ver com o último). De acordo com Nerlich, explicar a trajetória seguida por uma partícula livre observando que ela segue uma geodésica no espaço-tempo é uma ilustração de uma explicação geométrica em vez de causal. Uma teoria de explicação realmente satisfatória deve fornecer alguma resposta baseada em princípios para a questão de se A última seção mostra que, nas décadas de 1920 e 1930, houve algumas tentativas de desenvolver lógicas da descoberta. Mas, por várias décadas, a distinção de contexto ditou o que a filosofia da ciência deveria ser e como ela deveria proceder. todas as explicações por que devem ser causais (e de acordo com que noção de causal isso o é ou não), em vez de apenas assumir uma resposta afirmativa (ou negativa) para essa questão. Novamente, na medida em que existem formas não causais de explicação, a explicação permanecerá um tópico que é pelo menos um pouco independente da causalidade.

7.2 Explicação e outros objetivos epistêmicos

Noretta Koertge (1992) observou que, embora a literatura sobre explicação seja imensa, comparativamente pouca atenção tem sido dada, na construção dos vários modelos concorrentes de explicação, à questão de **para** o que eles devem ser usados ou qual é seu objetivo ou propósito (além de capturar "nossa" noção de explicação). Da mesma forma, os escritores sobre explicação nem sempre prestaram atenção adequada a como a própria explicação está conectada ou interage com (ou é distinta de) outros objetivos da investigação - por exemplo, qual é a conexão entre a bondade explicativa e outros objetivos frequentemente propostos para investigação tais como suporte de evidências, previsão, controle da natureza, simplicidade e assim por diante. Um resultado é que às vezes não é claro como avaliar a importância de nossos julgamentos intuitivos sobre a bondade de várias explicações ou determinar o que leva a que façamos um julgamento em vez de outro. Por exemplo, como observamos, a maioria das pessoas julga intuitivamente que não se pode explicar a altura de um mastro apelando para o comprimento de sua sombra.

Porém, um defensor decidido do modelo ND (HEMPEL, 1965, p. 353-4) pode muito bem perguntar por que deveríamos ficar tão impressionados com tais juízos intuitivos. Talvez nossa avaliação pré-analítica esteja confusa ou equivocada de alguma forma ou talvez reflita considerações meramente pragmáticas que não deveriam ter lugar na teoria da explicação. Uma maneira de responder a este ceticismo seria fornecer uma explicação sem petições de princípio sobre o que importante seria perdido ou ficaria de fora se não conseguíssemos distinguir entre as explicações dos comprimentos das sombras em termos de alturas dos mastros e "explicações" ocorrendo na direção oposta (observe que, na medida em que estamos apenas interessados em previsão, as duas inferências parecem estar no mesmo nível. "Sem petição de princípio" significa que não dizemos apenas que a altura causa a sombra e não vice-versa, mas que forneçamos alguma explicação adicional sobre em que consiste essa diferença e por qual razão a diferença é importante). Uma resposta possível apelaria para o objetivo epistêmico de ter informações relevantes para manipulação e controle; pode-se manipular o comprimento da sombra, entre outras coisas, manipulando a altura do mastro, mas não o contrário. Esta diferença é real, independentemente das intuições de alguém sobre a explicação

nos dois casos[30] .

Independentemente do que se pensa sobre esta resposta em particular, o ponto mais geral é que uma maneira de avaliar modelos concorrentes de explicação é focar menos, ou não apenas, em se eles capturam nossos julgamentos intuitivos e mais na questão de se e por qual razão os tipos de informações que eles exigem são valiosos (e atingíveis) e como essas informações se relacionam com outros objetivos que valorizamos na investigação.

Como outra ilustração, considere o modelo MC. Subjacente a esse modelo está, presumivelmente, algum juízo no sentido de que rastrear processos causais e suas interações é um objetivo digno de investigação. É claro que alguém pode tentar defender esse julgamento simplesmente alegando que a identificação das causas é um objetivo importante e que as teorias do processo causal fornecem a explicação correta da causa. Mas uma maneira mais esclarecedora e menos questionável de proceder seria perguntar como esse objetivo se relaciona com outros valores epistêmicos. Por exemplo, qual é a conexão entre o objetivo de identificar processos causais e construir teorias unificadas? Ou entre a identificação de processos causais e a descoberta de informações que são relevantes para a previsão ou para a manipulação e controle? Esses são os mesmos objetivos? Objetivos independentes, mas complementares? Objetivos concorrentes no sentido de que a satisfação de um pode tornar mais difícil satisfazer o outro? Obviamente, pode-se fazer perguntas semelhantes sobre o objetivo da unificação.

A necessidade de abordagens de explicação que relacionem esta noção mais adequadamente a outros conceitos e objetivos é particularmente saliente em conexão com o papel das leis na explicação, que é outro item na agenda para trabalhos futuros nesta área. O relato das leis, atualmente considerado como o mais promissor por muitos filósofos, é a teoria de Mill-Ramsey-Lewis (MRL). De acordo com essa teoria, as leis são aquelas generalizações que figuram como axiomas ou teoremas na sistematização dedutiva de nosso conhecimento empírico que atinge a melhor combinação de simplicidade e força (onde força tem a ver com a gama

---

[30] Essa assimetria está intimamente ligada ao fato de que o comprimento da sombra é o resultado de vários fatores diferentes que são independentemente variáveis; a altura do mastro, a orientação do mastro e a distância angular do sol acima do horizonte. Pode-se manipular o comprimento da sombra alterando os dois últimos fatores e isso não afetará a altura do mastro. Veja Hausman (1998) e Woodward (2003).

de verdades empíricas dedutíveis)[31]. É natural conectar esta concepção de leis com abordagens unificacionistas de explicação: se as leis são generalizações que desempenham um papel central na realização de sistematizações dedutivas simples (e, presumivelmente, unificadas), então, apelando para as leis na explicação, alcançamos a unificação explicativa - isso torna inteligível porque é desejável que as explicações invoquem leis[32]. Se uma teoria desse tipo pudesse funcionar, teríamos uma espécie de história integrada sobre leis e explicação que falta amplamente na teoria ND, uma história sobre o que são as leis que está diretamente conectada a uma ideia sobre o objetivo de explicação. É claro que permanecem problemas reais (alguns dos quais são discutidos acima) com a explicação unificacionista da explicação e, nesse caso, com a teoria das leis MRL[33], mas a explicação integrada que resultaria da junção dos dois, no entanto, pode ser tomado para ilustrar o tipo de coisa que devemos ter como objetivo.

7.3 Um único modelo de explicação?

Ainda outra questão geral diz respeito a até que ponto é possível construir um único modelo de explicação que se adapta a todas as áreas da ciência. É incontestável que a prática explicativa - o que é aceito como uma explicação, como os objetivos explicativos interagem com os outros, que tipo de informação explicativa é considerada alcançável, detectável, testável etc. – varia de maneiras significativas nas diferentes disciplinas. Não obstante, todos os modelos de explicação pesquisados acima são "universalistas" na aspiração – eles afirmam que um único modelo de explicação de "tamanho único" se encaixa em todas as áreas de investigação, na

---

[31] Para uma exposição da ideia básica, ver Lewis (1973a).

[32] Para uma observação semelhante, ver Psillos (2002). Observe, porém, que embora esse raciocínio forneça uma explicação de porque é **desejável** construir explicações apelando a leis (quando isso for possível), não está claro que produza o resultado de que explicação sempre **requeira** leis. Talvez haja generalizações que se unificam o suficiente para serem qualificadas como explicativas, de acordo com a abordagem unificacionista, mas não contam como leis de acordo com a abordagem do **MRL**. A versão de Kitcher do modelo unificacionista parece admitir isso como uma possibilidade.

[33] Veja Woodward (2003, p. 288-95, 358-73).

medida em que têm uma pretensão legítima de explicar. Embora a posição extrema de que a explicação na biologia ou história não tem nada em comum de interessante com a explicação na física pareça pouco atraente (e, em qualquer caso, atraiu pouco apoio), parece razoável esperar que mais esforço seja dedicado no futuro para desenvolver modelos de explicação que sejam mais sensíveis às diferenças disciplinares. Idealmente, esses modelos revelariam semelhanças entre as disciplinas, mas também deveriam nos permitir ver por que a prática explicativa varia tanto entre as diferentes disciplinas e o significado de tal variação. Por exemplo, como observado acima, biólogos, em contraste com físicos, muitas vezes descrevem seus objetivos explicativos como a descoberta de mecanismos, em vez da descoberta de leis. Embora seja concebível que essa diferença seja puramente terminológica, também vale a pena explorar a possibilidade de que haja uma história distinta a ser contada sobre o que é um mecanismo, como essa noção é compreendida pelos biólogos, e como as informações sobre os mecanismos contribuem para a explicação.

Um ponto intimamente relacionado é que pelo menos alguns dos modelos descritos acima impõem requisitos de explicação que podem ser satisfatórios em alguns domínios de investigação, mas são inatingíveis (em qualquer sentido praticamente interessante) em outros domínios ou, na medida em que podem ser alcançáveis, não têm relação perceptível com os objetivos geralmente aceitos de investigação nesses domínios. Por exemplo, observamos acima que muitos cientistas e filósofos sustentam que existem poucas ou nenhumas leis a serem descobertas na biologia e nas ciências sociais e comportamentais. Se é assim, modelos de explicação que atribuem um papel central às leis podem não ser muito esclarecedores sobre como a explicação funciona nessas disciplinas. Como outro exemplo, mesmo que suponhamos que a partição em classes de referência objetivamente homogêneas recomendadas pelo modelo RE é uma meta alcançável em conexão com certos fenômenos da mecânica quântica, pode ser que (como sugerido acima) simplesmente não seja uma meta que pode ser alcançada de uma forma não trivial em economia e sociologia, disciplinas nas quais inferência causal de estatísticas também figura com destaque. Em tais disciplinas, pode ser que partições adicionais estatisticamente relevantes de qualquer população ou subpopulação de interesse sejam virtualmente sempre possíveis, de modo que a atividade de localizar tais partições seja limitada apenas pelos custos de coleta de informações adicionais. Uma avaliação semelhante pode ser válida para a maioria das aplicações do modelo MC às ciências sociais.

Referências

ACHINSTEIN, P. **The Nature of Explanation**. New York: Oxford University Press, 1983.

ACHINSTEIN, P. **Evidence, Explanation, and Realism**: Essays in Philosophy of Science. New York: Oxford University Press, 2010.

BARNES, E. Explanatory Unification and the Problem of Asymmetry. **Philosophy of Science**, v. 59, p. 558-571, 1992.

BRAITHWAITE, R. **Scientific Explanation**. Cambridge: Cambridge University Press, 1953.BROMBERGER, S. Why Questions. *In:* COLODNY, R. (ed). **Mind and Cosmos**: Essays in Contemporary Science and Philosophy. Pittsburgh: University of Pittsburgh Press, 1966.

CARTWRIGHT, N. Causal Laws and Effective Strategies. **Noûs**, v. 13, p. 419-437, 1979.CARTWRIGHT, N. **How the Laws of Physics Lie**. Oxford: Clarendon Press, 1983.

CARTWRIGHT, N. From Causation to Explanation and Back. *In:* LEITER, B. (ed.). **The Future for Philosophy**. Oxford: Oxford University Press, 2004.

DE REGT, H.; DIEKS, D. A Contextual Account of Scientific Understanding. **Synthese**, v. 144, p. 137-170, 2005.

DOWE, P. **Physical Causation**. Cambridge: Cambridge University Press, 2000.

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

FRIEDMAN, M. Explanation and Scientific Understanding. **Journal of Philosophy**, v. 71, p. 5-19, 1974.

GARDINER, P. **The Nature of Historical Explanation**. Oxford: Oxford University Press, 1959.

GARDINER, P. (ed.). **Theories of History**. New York: The Free Press, 1952.

GOODMAN, N. **Fact, Fiction, and Forecast**. Cambridge: Harvard University Press, 1955.

GREENO, J. Evaluation of Statistical Hypotheses Using Information Transmitted. **Philosophy of Science**, v. 37, p. 279-293, 1970.

HALL, N. Two Concepts of Causation. *In:* COLLINS, J.; HALL, N.; Paul, L. (ed.). **Causation and Counterfactuals**. Cambridge: MIT Press, 2004, p. 225-276.

HAUSMAN, D. **Causal Asymmetries**. Cambridge: Cambridge University Press, 1998.

HAUSMAN, D.; WOODWARD, J. Independence, Invariance, and the Causal Markov Condition. **British Journal for the Philosophy of Science**, v. 50, p. 521-583, 1999.

HEMPEL, C. **Aspects of Scientific Explanation and Other Essays in the Philosophy of Science**. New York: Free Press, 1965a.

HEMPEL, C. Aspects of Scientific Explanation. 1965b *In:* HEMPEL, C. **Aspects of Scientific Explanation and Other Essays in the Philosophy of Science**. New York: Free Press, 1965a, p. 331-496.

HEMPEL, C. The Function of General Laws in History. **Journal of Philosophy**, v. 39, p. 35-48, 1942.

HEMPEL, C.; OPPENHEIM, P. Studies in the Logic of Explanation. **Philosophy of Science**, v. 15, p. 135-175, 1948.

JANZING, D.; MOOIJ, J.; ZHANG, K.; LEMEIRE, J.; ZSCHEISCHLER, J.; DANIUSIS, D.; STEUDEL, B.; SCHOLKOPF, B. Information-geometric Approach to Inferring Causal Directions. **Artificial Intelligence**, v.182-183, p. 1-31, 2012.

JEFFREY, R. Statistical Explanation vs. Statistical Inference. *In:* RESCHER, N. (ed.). **Essays in Honor of Carl G. Hempel**. Dordrecht: D. Reidel, 1969.

HAUSMAN, D. **Causal Asymmetries**. Cambridge: Cambridge University Press, 1998.

HITCHCOCK, C. Discussion: Salmon on Explanatory Relevance. **Philosophy of Science**, v. 62, p. 304-320, 1995.

KITCHER, P. Explanatory Unification and the Causal Structure of the World. *In:* Kitcher, P.; SALMON, W. **Scientific Explanation**. Minneapolis: University of Minnesota Press, 1989, p. 410-505.

KITCHER, P.; SALMON, W. Van Fraassen on Explanation. **Journal of Philosophy**, v. 84, p. 315-330, 1987.

KOERTGE, N. Explanation and Its Problems. **British Journal for the Philosophy of Science**, v. 43, p. 85-98, 1992.

KYBURG, H. Comment. **Philosophy of Science**, v. 32, p. 147-151, 1965.

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

LEWIS, D. Causation. **Journal of Philosophy**, v. 70, p. 556-567.

LEWIS, D. **Philosophical Papers**. Oxford: Oxford University Press, 1986. v. 2.

LEWIS, D. Causation as Influence. **Journal of Philosophy**, v.97, p. 182-197, 2000.

LIEN, Y.; CHENG, P. Distinguishing genuine from spurious causes: a coherence hypothesis. **Cognitive Psychology**, v. 40, p. 87-137, 2000.

LOMBROZO, T. Causal-Explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions. **Cognitive Psychology**, v. 61, p. 303-332, 2010.

MITCHELL, S. Pragmatic Laws' PSA 96. **Philosophy of Science**, supplement, v. 64, n. 4, p. 468-479, 1997.

MORRISON, M. **Unifying Scientific Theories**. Cambridge: Cambridge University, Press, 2000.

NAGEL, E. **The Structure of Science**: Problems in the Logic of Scientific Explanation. New York: Harcourt, Brace and World, 1961.

NERLICH, G. What Can Geometry Explain? **British Journal for the Philosophy of Science**, v. 30, p. 69-83, 1979.

NORTON, J. A Material Dissolution of the Problem of Induction. **Synthese**, v. 191, p. 671-690, 2014.

PEARL, J. **Causality**: Models, Reasoning and Inference, Cambridge: Cambridge University, 2000.

PITT, J. (ed.). **Theories of Explanation**. New York: Oxford University Press, 1988.

POPPER, K. **The Logic of Scientific Discovery**. London: Hutchinson, 1959.

PSILLOS, S. **Causal Asymmetries**. Stocksfield: Acumen Publishing, 2002.

RAILTON, P. A Deductive-Nomological Model of Probabilistic Explanation. **Philosophy of Science**, v. 45, p. 206-226, 1978.

RAILTON, P. Probability, Explanation, and Information. **Synthese**, v. 48, p. 233-256, 1981.

RUBEN, D. (ed.). **Explanation.** Oxford: Oxford University Press, 1993.

SALMON, W. **Four Decades of Scientific Explanation.** Minneapolis: University of Minnesota Press, 1989.

SALMON, W. Statistical Explanation. *In:* SALMON, W. (ed.). **Statistical Explanation and Statistical Relevance.** Pittsburgh: University of Pittsburgh Press, 1971a, p. 29-87.

SALMON, W. **Four Decades of Scientific Explanation.** Minneapolis: University of Minnesota Press, 1989.

SALMON, W. **Scientific Explanation and the Causal Structure of the World.** Princeton: Princeton University Press, 1984.

SALMON, W. Causality Without Counterfactuals. **Philosophy of Science**, v. 61, p. 297-312, 1994.

SALMON, W. Causality and Explanation: A Reply to Two Critiques. **Philosophy of Science**, v. 64, p. 461-477, 1997.

SALMON, W. (ed.). **Statistical Explanation and Statistical Relevance**. Pittsburgh: University of Pittsburgh Press, 1971b.

SALMON, W.; KITCHER, P. (eds.). **Minnesota Studies in the Philosophy of Science, Scientific Explanation**. Minneapolis: University of Minnesota Press, 1989. v. 13.

SCHAFFER, J. Causation by Disconnection. **Philosophy of Science**, v. 67, p. 285-300, 2000.

SCRIVEN, M. Truisms as the Grounds of Historical Explanations. *In:* GARDINER, P. **The Nature of Historical Explanation**. Oxford: Oxford University Press, 1959.

SCRIVEN, M. Explanations, Predictions, and Laws. *In:* FEIGL, H.; MAXWELL, G. (ed.). **Scientific Explanation, Space, and Time**. Minneapolis: University of Minnesota Press, 1962, p.170-230. v. 3.

SOBER, E. Equilibrium Explanation. **Philosophical Studies**, v. 43, p. 201-210, 1983.

SOBER, E. The Multiple Realizability Argument Against Reductionism. **Philosophy of Science**, v. 66, p. 542-564, 1999.

SOBER, E. Two Uses of Unification. *In:* STADLER, F. (ed.). **The Vienna Circle and Logical Empiricism**: Re-Evaluation and Future Perspectives. Viena: Institute Vienna Circle Yearbook, 2003, p. 205-216.

SPIRTES, P.; GLYMOUR, C.; SCHEINES, R. **Causation, Prediction and Search**. Cambridge: MIT Press, 1983.

VAN FRAASSEN, B. C. **The Scientific Image**. Oxford: Clarendon Press, 1980.

WOODWARD, J. The Causal/Mechanical Model of Explanation. In: KITCHER, P.; SALMON, W. (ed.). **Scientific Explanation**. Minnesota: University of Minnesota Press, 1989, p. 357-383. v. 13

WOODWARD, J. Explanation and Invariance in the Special Sciences. **British Journal for the Philosophy of Science**, v. 51, p. 197–254, 2000.

WOODWARD, J. **Making Things Happen:** A Theory of Causal Explanation. Oxford: Oxford University Press, 2003.

WOODWARD, J. Sensitive and Insensitive Causation. **The Philosophical Review**, v. 115, p. 1-50, 2006.

(II) Descoberta Científica*

Autor: Jutta Schickore
Tradução: Jessica Suellen Oliveira Costa
Revisão: Mariana Vitti Rodrigues

Descoberta científica é o processo ou o produto da investigação científica bem sucedida. Objetos da descoberta podem ser coisas, eventos, processos, causas e propriedades, bem como teorias, hipóteses e suas características (seu poder explanatório, por exemplo). A maioria das discussões filosóficas sobre descobertas científicas tem seu foco na geração de novas hipóteses que se enquadram ou explicam determinados conjuntos de dados ou possibilitam a derivação de consequências testáveis. Discussões filosóficas sobre a descoberta científica têm sido bastante intrincadas e complexas, pois o termo "descoberta" é usado de diferentes maneiras, tanto para referir-se ao resultado quanto ao procedimento de investigação. Em sentido mais estreito, o termo "descoberta" refere-se ao suposto "momento eureka" sobre um novo *insight*. Já em sentido mais amplo, "descoberta"

---

* Schickore, J. Scientific Discovery. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Summer Edition. Stanford, CA: The Metaphysics Research Lab, 2018. Disponível em: https://plato.stanford.edu/archives/sum2018/entries/scientific-discovery/. Acesso em: 18 set. 2021.

The following IE the translation of the entry on Scientific Explanation by Jutta Schickore, 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/sum2018/entries/scientific-discovery/. 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 IE located at https://plato.stanford.edu/entries/scientific-discovery/. 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.

é um sinônimo para “empreendimento científico bem sucedido”. Alguns debates filosóficos a respeito da natureza da descoberta científica refletem essas variações terminológicas.

Questões filosóficas, relacionadas à descoberta científica, surgem sobre a natureza da criatividade humana, especificamente sobre se o “momento eureka” pode ser analisado e se existem regras (algoritmos, orientações ou heurísticas) segundo as quais um novo insight pode ser provocado. Questões filosóficas também surgem sobre heurística racional, características de hipóteses dignas de articulação e testes e, em um metanível, sobre a natureza e o escopo da própria reflexão filosófica. Este ensaio descreve a emergência e o desenvolvimento do problema filosófico da descoberta científica, apresenta diferentes abordagens filosóficas para entender a descoberta científica e indica problemas metafilosóficos em torno desse debate.

1. Introdução

A reflexão filosófica sobre a descoberta científica ocorreu em diferentes fases. Antes da década de 1930, filósofos estavam interessados principalmente em descobertas no sentido mais amplo do termo, isto é, com a análise da investigação científica bem sucedida como um todo. Discussões filosóficas focavam em saber se havia algum padrão discernível na produção de novos conhecimentos. Como o conceito de descoberta não tinha um significado específico e era usado em um sentido mais amplo, quase todos os tratados dos séculos XVII e XVIII sobre o método científico podem ser potencialmente considerados como contribuições iniciais para reflexões sobre descoberta científica. No decorrer do século XIX, quando a filosofia da ciência e a ciência se tornaram dois empreendimentos distintos, o termo “descoberta” passou a ser um termo técnico nas discussões filosóficas. Diferentes elementos da investigação científica foram especificados. Mais importante, a geração de novos conhecimentos foi clara e explicitamente diferenciada de sua validação e, dessa forma, as condições para o sentido mais estreito de descoberta, como ato ou processo de concepção de novas ideias, emergiram.

A próxima fase na discussão sobre descoberta científica começou com a introdução da chamada “distinção de contexto”, a distinção entre o “contexto de descoberta” e o “contexto de justificação”. Foi ainda discutido que o ato de conceber uma nova ideia é um processo não racional, um salto de insight que não pode ser capturado em instruções específicas. Justificação, ao contrário, é um processo sistemático de estabelecimento de critérios avaliativos para reivindicações de

conhecimento. Os defensores da distinção de contexto argumentaram que a filosofia da ciência se interessava apenas com o contexto da justificação. A suposição subjacente a esse argumento é de que a filosofia é um projeto normativo; determina as normas para a prática científica. Dadas essas suposições, apenas a justificação de ideias pode ser objeto da análise filosófica (normativa), não a sua geração. Descoberta, por outro lado, pode ser apenas um tópico de estudo empírico. Por definição, o estudo da descoberta está fora do escopo da própria filosofia da ciência.

A introdução da distinção de contexto e da distinção disciplinar, que estava ligada a ela, geraram disputas metafilosóficas. Durante muito tempo, debates filosóficos sobre descoberta foram norteados pela noção de que análises filosóficas e empíricas são mutuamente exclusivas. Muitos filósofos insistiram, assim como seus antecessores antes da década de 1930, que as tarefas dos filósofos incluíssem a análise das práticas científicas vigentes e que recursos científicos fossem utilizados para direcionar problemas filosóficos. Eles também sustentaram que é uma tarefa legítima para a filosofia da ciência o desenvolvimento de uma teoria de heurística ou de resolução de problemas. Mas essa posição foi minoritária durante grande parte da filosofia da ciência do século XX. Filósofos da descoberta foram, assim, forçados a demonstrar que a descoberta científica foi, de fato, uma parte legítima da filosofia da ciência. Reflexões filosóficas sobre a natureza da descoberta científica tiveram de ser reforçadas por argumentos metafilosóficos sobre a natureza e o escopo da filosofia da ciência.

Hoje, no entanto, existe um amplo consenso de que a filosofia e a pesquisa empírica não são mutuamente exclusivas. Estudos empíricos de autênticas descobertas científicas não apenas informam o pensamento filosófico sobre mecanismos estruturais e cognitivos da descoberta, mas pesquisas em psicologia, ciência cognitiva, inteligência artificial e outros campos relacionados tornaram-se parte integral da análise filosófica dos processos e das condições para a geração de novos conhecimentos.

2. A investigação científica como descoberta

Antes do século XIX, o termo “descoberta” se referia comumente ao produto de investigação bem sucedida. O termo “descoberta” foi usado amplamente para referir-se a novos achados, como uma nova cura, o aperfeiçoamento de um instrumento ou um novo método para mensurar longitude. Muitos filósofos naturais

e experimentais, especialmente Bacon, Descartes e Newton, expuseram propostas de métodos científicos para alcançar novos conhecimentos. Essas propostas não foram explicitamente rotuladas "métodos de descoberta", mas as propostas gerais de métodos científicos são, no entanto, relevantes para debates filosóficos atuais sobre descoberta científica. São relevantes pois filósofos da ciência têm frequentemente apresentado as teorias do método científico do século XVII como uma categoria contrastante às filosofias de descoberta atuais. A característica distintiva das propostas de método científico dos séculos XVII e XVIII é de que os métodos são considerados como tendo força probatória (NICKLES, 1985). Isso quer dizer que aquelas propostas de método científico funcionam como guia para a aquisição de novos conhecimentos e, ao mesmo tempo, como validações dos conhecimentos assim obtidos (LAUDAN, 1980; SCHAFFNER, 1993, cap. 2).

O relato de Bacon sobre o "novo método", como apresentado no *Novum Organum*, é um exemplo proeminente. A obra de Bacon mostrou qual a melhor maneira de chegar ao conhecimento sobre as "naturezas da forma" (as propriedades mais gerais da matéria) através de uma investigação sistemática de naturezas fenomenais. Bacon descreveu primeiramente como coletar e organizar fenômenos naturais e fatos experimentais em tabelas, como avaliar essas listas e como refinar os resultados iniciais com a ajuda de outros experimentos. Por meio dessas etapas, o investigador chegaria a conclusões sobre a "natureza da forma", que produz naturezas fenomenais particulares. A questão é que, para Bacon, os procedimentos de elaborar e avaliar tabelas bem como conduzir experimentos de acordo com o *Novum Organum*, levam ao conhecimento seguro. Portanto, os procedimentos têm "força probatória".

Similarmente, o objetivo de Newton em *Philosophiae Naturalis Principia Mathematica* era apresentar um método para dedução de proposições de fenômenos, de modo que essas proposições se tornassem "mais seguras" que as proposições que são asseguradas pela dedução de suas consequências testáveis (SMITH, 2002). Newton não assumiu que esse procedimento pudesse levar a certeza absoluta. Só se poderia obter certeza moral para as proposições garantidas deste modo. A questão para os atuais filósofos da ciência é de que essas abordagens são teorias gerativas do método científico. As teorias gerativas do método científico assumem que proposições só podem ser estabelecidas e asseguradas mostrando que elas se seguem de fenômenos observados e experimentalmente produzidos. Em contraste, teorias não gerativas do método científico, como proposto por Huygens, assumem que as proposições devem ser estabelecidas comparando suas consequências

com os fenômenos observados e experimentalmente produzidos. Na filosofia da ciência do século XX, essa abordagem é frequentemente caracterizada como "consequencialista" (LAUDAN, 1980; NICKLES, 1985).

Filósofos contemporâneos da ciência têm usado esboços históricos como esses para reconstruir a pré-história dos atuais debates filosóficos sobre a descoberta científica. O argumento é que a descoberta científica passou a ser um problema para a filosofia da ciência no século XIX, quando as teorias consequencialistas do método científico se tornaram mais difundidas. Quando as teorias consequencialistas estavam em ascensão, os dois processos de concepção e de validação de uma ideia ou hipótese tornaram-se distintos e a consideração de que o mérito de uma nova ideia não depende do modo no qual ela foi alcançada, se tornou amplamente aceita.

3. Elementos de descoberta

No decorrer do século XIX, o ato de ter um *insight*, o suposto "momento eureka", foi separado dos processos de articular, desenvolver e testar o novo insight. A discussão filosófica focalizou a questão de saber se, em que medida, regras poderiam ser concebidas para orientar cada um desses processos. O trabalho de William Whewell, especialmente os dois volumes de *Philosophy of the Inductive Sciences*, de 1840, é uma contribuição importante para os debates filosóficos sobre descoberta científica, justamente porque ele separou claramente o momento criativo ou, como ele chamou, "pensamento feliz" (*happy thought*), de outros elementos da investigação científica. Para Whewell, descoberta compreende todos os três elementos: o pensamento feliz, a articulação e o desenvolvimento desse pensamento, e seu teste ou verificação. Na maioria das propostas subsequentes sobre descoberta, entretanto, o escopo do termo "descoberta" limita-se ou ao primeiro desses elementos, o "pensamento feliz", ou aos dois primeiros desses elementos, o pensamento feliz e sua articulação. De fato, muitas das controvérsias no século XX sobre a possibilidade de uma filosofia da descoberta podem ser entendidas no contexto da divergência sobre se o processo de descoberta inclui ou não a articulação e o desenvolvimento de um novo pensamento.

A seção anterior mostra que pesquisadores, como Bacon e Newton, pretendiam desenvolver metodologias de investigação científica. Eles propuseram "novos métodos", ou "regras de raciocínio", que direcionam a produção de certas proposições a partir de fenômenos observados e experimentais. Em contraste,

Whewell estava explicitamente interessado em desenvolver uma filosofia da descoberta. Sua proposta era em parte uma descrição da composição psicológica do descobridor. Por exemplo, ele sustentou que apenas os gênios poderiam ter esses pensamentos felizes, que são essenciais para a descoberta. Em parte, sua proposta era sobre os métodos nos quais os pensamentos felizes são integrados ao sistema de conhecimento. De acordo com Whewell, o passo inicial em toda descoberta é o que ele chamou de "algum pensamento feliz, no qual não podemos traçar a origem; alguma projeção afortunada do intelecto, elevando-se acima de todas as regras. Nenhuma máxima pode ser dada que inevitavelmente conduza à descoberta" (WHEWELL, 1996, p. 186). Uma "arte da descoberta", no sentido de uma habilidade que pode ser ensinada e aprendida, não existe de acordo com Whewell. O pensamento feliz se constrói sobre fatos conhecidos, mas, segundo Whewell, é impossível prescrever um método para ter pensamentos felizes.

Nesse sentido, pensamentos felizes são acidentais. Porém, em um sentido importante, descobertas científicas **não** são acidentais. O pensamento feliz não é uma adivinhação grosseira. Somente a pessoa na qual a mente esteja preparada para ver objetos realmente os notará. A "condição prévia do intelecto, e não um fato único, é realmente a causa principal e peculiar do sucesso. O fato é apenas a ocasião em que o mecanismo da descoberta entra em jogo, mais cedo ou mais tarde. Isso é, como já disse em outro lugar, apenas a faísca que dispara uma arma já carregada e apontada; e há pouca propriedade em falar de um acidente como causa na qual a bala atinge seu alvo" (WHEWELL, 1996, p. 189).

Contudo, ter um pensamento feliz ainda não é uma descoberta. O segundo elemento de uma descoberta científica consiste em unir, "coligar" (*colligating*), como Whewell o chamou, um conjunto de fatos através da conjunção desses fatos por meio de uma concepção geral. A coligação não apenas produz algo novo, mas também mostra os fatos previamente conhecidos sob uma nova luz. Mais precisamente, a coligação funciona em ambos os sentidos, a partir de fatos e também a partir das ideias que unem os fatos. A coligação é um processo estendido. Ela envolve, por um lado, a especificação de fatos através de observação sistemática, mensurações e experimentos e, por outro lado, o esclarecimento de ideias por meio de exposição das definições e axiomas que estão tacitamente implícitos nessas ideias. Esse processo é iterativo. Os cientistas vão e voltam entre unir os fatos, esclarecer a ideia, tornar os fatos mais exatos, e assim por diante.

A última etapa da descoberta é a verificação da coligação que envolve o pensamento feliz. Isso quer dizer, primeira e principalmente, que o resultado da

coligação deve ser o suficiente para explicar os dados em questão. A verificação também envolve o julgamento do poder preditivo, da simplicidade e da "consiliência" (*consilience*) do resultado da coligação. A "consiliência" refere-se a uma gama maior de generalidade (aplicabilidade mais ampla) da teoria (o pensamento feliz articulado e esclarecido) que a coligação verdadeira produziu. A proposta de Whewell sobre a descoberta não é concebida como um sistema dedutivista. É essencial que o resultado da coligação seja passível de ser inferido a partir de dados anteriores a qualquer teste (SNYDER, 1997).

A teoria da descoberta de Whewell é significativa para o debate filosófico sobre descoberta científica, pois separa claramente três elementos: o pensamento feliz não analisável ou "momento eureka"; o processo da coligação no qual inclui-se o esclarecimento e a explicação de fatos e ideias; e a verificação do resultado da coligação. Sua posição, de que a filosofia da descoberta não pode prescrever como conceber pensamentos felizes, tem sido um elemento-chave na reflexão filosófica do século XX sobre a descoberta. Notavelmente, contudo, a concepção de Whewell sobre descoberta não compreende apenas os pensamentos felizes, mas também os processos pelos quais os pensamentos felizes são integrados em um dado sistema de conhecimento. Os procedimentos de articulação e de teste são **ambos** analisáveis de acordo com Whewell, e sua concepção de coligação e verificação serve como um guia sobre como o descobridor deve proceder. Uma coligação, se realizada adequadamente, possui **tal** força de justificação. Da mesma maneira, o processo de verificação é uma parte integral da descoberta e também possui força de justificação. A concepção de Whewell sobre a verificação compreende, portanto, elementos de métodos gerativos e consequenciais de investigação. Para verificar uma hipótese, o investigador precisa mostrar que ela considera os fatos conhecidos, prediz novos fenômenos previamente não observados e que pode explicar e prever fenômenos que são explicados e previstos por uma hipótese obtida através de um pensamento feliz a partir de uma coligação (DUCASSE, 1951).

A conceitualização de Whewell sobre a descoberta científica oferece uma estrutura útil para mapear os debates filosóficos sobre a descoberta e identificar os principais problemas de interesse nos debates filosóficos recentes. Em primeiro lugar, quase todos os filósofos recentes operam com uma noção de descoberta mais restrita que a de Whewell. Na concepção mais restrita, o que Whewell chamou de "verificação" não faz parte da descoberta propriamente dita. Segundo, até o final do século XX, havia um amplo consenso de que o momento eureka, interpretado de maneira restrita, não é analisável, sendo até mesmo um ato misterioso de insight.

As principais divergências sobre a questão de se o processo de desenvolvimento de uma hipótese (a "coligação", nos termos de Whewell) é, ou não, uma parte da própria descoberta, e, se for, se e como esse processo é guiado por regras. Filósofos também discordaram sobre o problema de saber se explicar essas regras seria uma tarefa filosófica. Nas décadas recentes, a atenção filosófica mudou para o momento eureka. Com base nos recursos da ciência cognitiva, neurociência, pesquisa computacional e psicologia ambiental e social, filósofos procuraram desmistificar os processos cognitivos envolvidos na geração de novas ideias.

4. Lógicas da descoberta

No início do século XX, a concepção de que a descoberta era, ou que pelo menos envolvia crucialmente, um ato criativo não analisável de um gênio talentoso, era preponderante, mas não aceita por unanimidade. Concepções alternativas da descoberta enfatizam que a descoberta é um processo estendido, isto é, que o processo de descoberta inclui processos de raciocínio nos quais um novo insight é articulado e desenvolvido. Além disso, assumiu-se que há um aspecto formal e sistemático nesse raciocínio. Embora o raciocínio envolvido não prossiga de acordo com os princípios da lógica demonstrativa, ele é suficientemente sistemático para merecer o rótulo de "lógico". Os defensores dessa perspectiva argumentaram que a lógica tradicional, aristotélica, é um modelo inadequado de descoberta científica, pois interpreta erroneamente o processo de geração de conhecimento de maneira tão grosseira quanto a noção de "pensamento feliz". Nessa abordagem, o termo "lógica" é usado em sentido amplo. É tarefa da lógica da descoberta extrair e fornecer uma representação esquemática das estratégias de raciocínio utilizadas em episódios de investigação científica bem sucedida. As lógicas da descoberta, no início do século XX, podem ser mais bem descritas como teorias de operações mentais envolvidas na geração de conhecimento. Entre essas operações mentais, estão a classificação, a determinação do que é relevante para uma investigação e as condições de comunicação do significado. Argumenta-se que esses aspectos da descoberta científica são representados de maneira insuficiente pela lógica tradicional, ou não são representados (SCHILLER, 1917, p. 236-7).

Os filósofos defensores dessa abordagem concordam que a lógica da descoberta deve ser caracterizada como um conjunto de princípios heurísticos, ao invés de um processo de empregar a lógica indutiva ou dedutiva a um conjunto de

proposições. Esses princípios heurísticos não são entendidos como uma forma de mostrar o caminho para o conhecimento seguro. Os princípios heurísticos são sugestivos ao invés de serem demonstrativos (CARMICHAEL, 1922, 1930). Uma recorrente característica nas propostas de estratégias de raciocínio que possibilitam novas ideias, é o raciocínio analógico (SCHILLER, 1917; BENJAMIN, 1934). No século XX, o raciocínio analógico é amplamente reconhecido como uma forma frutífera de raciocínio que não pode ser reduzida a inferências indutivas ou dedutivas (*vide* seção 9.2). Entretanto, essas abordagens da lógica da descoberta permaneceram dispersas e provisórias naquele momento, e as tentativas de desenvolver processos heurísticos que orientam a descoberta de forma mais sistemática foram ofuscadas pelo avanço da distinção entre os contextos de descoberta e de justificação.

5. Distinção entre contexto de descoberta e contexto de justificação

A distinção entre "contexto de descoberta" e "contexto de justificação" dominou e moldou as discussões sobre descoberta na filosofia da ciência do século XX. A distinção de contexto marca a distinção entre a geração de uma nova ideia ou hipótese e sua defesa (teste, verificação). Como as seções anteriores mostraram, a distinção entre diferentes características da investigação científica tem uma longa história, mas na filosofia da ciência, essa distinção se tornou relevante na primeira metade do século XX. No decorrer das discussões que se seguiram sobre a descoberta científica, a distinção entre diferentes características da investigação científica tornou-se um poderoso critério de demarcação. Os limites entre o contexto da descoberta (os processos de pensamento *de facto*) e o contexto da justificativa (a defesa *de jure* da correção desses pensamentos) foram agora entendidos como determinantes do escopo da filosofia da ciência. A suposição subjacente é a de que a filosofia da ciência é um empreendimento normativo. Os defensores da distinção de contexto argumentam que a geração de uma nova ideia é um processo intuitivo, não racional; que não pode estar sujeita à análises normativas. Logo, o estudo do pensamento genuíno dos cientistas, só pode ser objeto da psicologia, sociologia e outras ciências empíricas. A filosofia da ciência, ao contrário, está exclusivamente interessada no contexto da justificativa.

Os termos "contexto de descoberta" e "contexto de justificação" são frequentemente associados à obra de Hans Reichenbach. A concepção original de Reichenbach sobre a distinção de contexto, entretanto, é bastante complexa (HOWARD,

2006; RICHARDSON, 2006). Não é fácil mapear a distinção disciplinar mencionada acima porque, para Reichenbach, a própria filosofia da ciência é parcialmente descritiva. Reichenbach sustenta que a filosofia da ciência inclui uma descrição do conhecimento como ele realmente é. A filosofia da ciência descritiva reconstrói os processos de pensamento dos cientistas de modo que possa realizar análises lógicas sobre eles e, dessa forma, fornece as bases para a avaliação desses pensamentos (REICHENBACH, 1938). A descoberta, por outro lado, é objeto de estudo empírico, psicológico e sociológico. De acordo com Reichenbach, o estudo empírico de descobertas mostra que os processos de descoberta correspondem frequentemente ao princípio da indução, mas isso é simplesmente um fato psicológico (Reichenbach, 1938: 403).

Embora os termos "contexto de descoberta" e "contexto de justificação" sejam amplamente utilizados, houve bastante discussão sobre como a distinção deve ser traçada e qual o seu significado filosófico (KORDIG, 1978; GUTTING, 1980; ZAHAR, 1983; LEPLIN, 1987; HOYNINGEN-HUENE, 1987; WEBER, 2005; SCHICKORE; STEINLE, 2006). Comumente, a distinção é interpretada como uma distinção entre o processo de concepção de uma teoria e sua validação, isto é, a determinação de suporte epistêmico da teoria. Essa versão da distinção não é necessariamente interpretada como uma distinção temporal. Em outras palavras, geralmente não se assume que a teoria seja primeiramente desenvolvida em sua totalidade e depois validada. Em vez disso, a concepção e a validação são duas abordagens epistêmicas diferentes da teoria: o esforço de articular, embasar e desenvolver seu potencial e o esforço de avaliar seu valor epistêmico. Dentro da perspectiva da distinção de contexto, existem duas maneiras principais de conceituar o processo de conceber uma teoria. A primeira opção é caracterizar a geração de novos conhecimentos como um ato irracional, uma misteriosa intuição criativa, um "momento eureka". A segunda opção é conceitualizar a geração de novos conhecimentos como um processo estendido que **inclui** um ato criativo, bem como algum processo de articulação e desenvolvimento da ideia criativa.

Ambas as considerações sobre geração de conhecimento serviram como ponto de partida para argumentos contra a possibilidade de uma filosofia da descoberta. De acordo com a primeira opção, os filósofos argumentaram que não é possível prescrever um método lógico que produza novas ideias, nem é possível reconstruir logicamente o processo de descoberta. Apenas o processo de teste é passível de investigação lógica. Essa objeção às filosofias da descoberta foi chamada de "objeção da máquina de descoberta" (CURD, 1980, p. 207). Ela é geralmente

associada com a obra *Logic of Scientific Discovery*, de Karl Popper.

> O estágio inicial, o ato de conceber ou inventar uma teoria, parece-me não exigir análise lógica, nem ser suscetível a ela. A questão de saber como uma nova ideia ocorre ao homem – seja sobre um tema musical, um conflito dramático ou uma teoria científica – pode ser de grande interesse para a psicologia empírica; mas isso é irrelevante para a análise lógica do conhecimento científico. Este último não é sobre *questões de fato* (*quid facti?* de Kant), mas apenas sobre *questões de justificação ou validade* (*quid juris?* de Kant). Suas questões são do seguinte tipo. Um enunciado pode ser justificado? E se sim, como? Ele é testável? É logicamente dependente de outros enunciados? Ou possivelmente os contradiga? [...] Deste modo, devo distinguir nitidamente entre o processo de conceber uma nova ideia e os métodos e resultados de examiná-la logicamente. Quanto à tarefa da lógica do conhecimento – em oposição à psicologia do conhecimento – devo proceder sobre a suposição de que ela consiste unicamente em investigar os métodos empregados nos testes sistemáticos aos quais toda nova ideia deve ser submetida se quiser ser seriamente cogitada (POPPER, 2002, p. 7-8).

Em respeito à segunda forma de conceitualização da geração de conhecimento, muitos filósofos argumentam, de modo semelhante, que uma lógica da descoberta não pode ser desenvolvida, pois o processo de descoberta envolve um processo irracional e intuitivo, que não pode ser examinado logicamente. Outros filósofos são contrários à filosofia da descoberta, embora reconheçam explicitamente que a descoberta é um processo estendido e fundamentado. Eles apresentam uma objeção metafilosófica, argumentando que uma teoria sobre a articulação e desenvolvimento de ideias não é uma teoria filosófica, mas sim, uma teoria psicológica.

O impacto da distinção de contexto nos estudos sobre a descoberta científica e na filosofia da ciência em geral pode dificilmente ser superestimado. A ideia de que o processo da descoberta (em qualquer de suas versões) está fora do escopo

da própria filosofia da ciência, foi amplamente compartilhada entre os filósofos da ciência durante a maior parte do século XX e ainda é defendida por muitos. A última seção mostra que, nas décadas de 1920 e 1930, houve algumas tentativas de desenvolver lógicas da descoberta. Mas, por várias décadas, a distinção de contexto ditou o que a filosofia da ciência deveria ser e como ela deveria proceder. A ideia dominante era que as teorias das operações mentais ou heurísticas não tinham lugar na filosofia da ciência. Por esse motivo, a descoberta não era um tópico legítimo para a filosofia da ciência. A noção ampla de descoberta é principalmente empregada em considerações sociológicas sobre a prática científica. Nesta perspectiva, "descoberta" é entendida como um rótulo retrospectivo, que é atribuído como um sinal de realização para alguns empreendimentos científicos. As teorias sociológicas reconhecem que a descoberta é uma conquista coletiva e o resultado de um processo de negociação por meio do qual "histórias de descoberta" são construídas e que certas reivindicações de conhecimento recebem o *status* de descoberta (Brannigan, 1981; Schaffer, 1986, 1994). Até o último terço do século XX, havia poucas tentativas de desafiar a distinção disciplinar ligadas à distinção de contexto. Somente na década de 1970, o interesse em abordagens filosóficas sobre a descoberta começou a aumentar. Mas a distinção de contexto permaneceu como um desafio para as filosofias da descoberta.

Existem três correntes principais de resposta à diferença disciplinar ligada à distinção de contexto. Cada uma dessas correntes de resposta inauguram uma perspectiva filosófica sobre a descoberta. Cada uma dessas correntes parte do pressuposto de que a filosofia da ciência pode legitimamente incluir alguma forma de análise de padrões *reais* de raciocínio, bem como informação das ciências empíricas, como a ciência cognitiva, psicologia e sociologia. Todas essas respostas rejeitam a ideia de que a descoberta não é nada além de um evento místico. A descoberta é concebida como um processo de raciocínio analisável, não apenas como um salto criativo pelo qual novas ideias surgem para serem totalmente formadas. Todas essas respostas concordam que os procedimentos e métodos para se chegar a novas hipóteses e ideias não garantem que a hipótese ou ideia, assim formada, seja necessariamente a melhor ou a correta. No entanto, é tarefa da filosofia da ciência fornecer regras para aprimorar esse processo. Todas essas respostas podem ser descritas como teorias de resolução de problemas, cujo objetivo final é tornar mais eficiente a geração de novas ideias e teorias.

As diferentes abordagens da descoberta científica, porém, empregam diferentes terminologias. Em particular, o termo "lógica" da descoberta é usado

algumas vezes em um sentido estrito e em outras vezes, compreendido em sentido amplo. No sentido estrito, "lógica" da descoberta deve ser entendida como referência à um conjunto de regras formais geralmente empregadas, pelas quais novas ideias podem ser mecanicamente derivadas a partir de dados existentes. No sentido amplo, "lógica" da descoberta refere-se à representação esquemática dos procedimentos de raciocínio. "Lógica" é apenas outro termo para "racional". Além disso, enquanto cada uma dessas respostas combina análises filosóficas da descoberta científica com pesquisas empíricas sobre a cognição humana de fato, diferentes conjuntos de recursos são mobilizados, desde a pesquisa em Inteligência Artificial (IA) e ciência cognitiva, até estudos históricos sobre procedimentos de resolução de problemas. Ademais, as respostas analisam o processo de investigação científica de maneira diferente. Frequentemente, a investigação científica é considerada como tendo dois aspectos, a saber, a geração e a validação de novas ideias. Entretanto, algumas vezes, a investigação científica é considerada como tendo três aspectos, a saber, geração, busca ou articulação, e validação de conhecimento. Neste último, o rótulo "descoberta" é usado algumas vezes para referir-se apenas à geração e, outras vezes, à geração e a busca.

A primeira resposta ao desafio da distinção de contexto baseia-se em uma ampla compreensão do termo "lógica" para argumentar que não podemos deixar de admitir uma lógica geral e neutra em relação ao seu domínio, se não quisermos assumir que o sucesso da ciência é um milagre (Jantzen, 2016) e que uma lógica da descoberta científica pode ser desenvolvida (seção 6). A segunda resposta, baseando-se no entendimento estrito do termo "lógica", concebe que não há *lógica* da descoberta, isto é, não há nenhum algoritmo que gere novos conhecimentos. Filósofos que adotam essa abordagem argumentam que o processo de descoberta segue um padrão identificável e analisável (seção 7). Outros, argumentam que a descoberta é governada por uma **metodologia**. A metodologia da descoberta é um tópico legítimo para a análise filosófica (seção 8). Todas essas respostas pressupõem que há mais elementos na descoberta do que um momento eureka. A descoberta compreende processos de articulação e desenvolvimento do pensamento criativo. Esses são os processos que podem ser examinados com as ferramentas de análise filosófica. A terceira resposta ao desafio da distinção de contexto também pressupõe que a descoberta é, ou que pelo menos envolve, um ato criativo. Mas ao contrário das duas primeiras respostas, ela se interessa com o próprio ato criativo. Filósofos que adotam essa abordagem argumentam que a criatividade científica é passível de análise filosófica (seção 9).

6. A lógica da descoberta depois da distinção de contexto

A primeira resposta ao desafio da distinção de contexto é argumentar que a descoberta é um tópico para a filosofia da ciência pois, afinal, ela é um processo lógico. Os defensores dessa abordagem da lógica da descoberta geralmente aceitam a distinção principal entre os processos de conceber e de testar uma hipótese. Eles também concordam que é impossível desenvolver um manual que forneça um procedimento formal e mecânico por meio dos quais conceitos inovadores ou hipóteses possam ser derivados: Não existe uma máquina de descoberta. Mas eles rejeitam a ideia de que o processo de conceber uma teoria é um ato criativo, uma suposição misteriosa, um palpite, um processo mais ou menos instantâneo e aleatório. Em vez disso, eles insistem que ambos, conceber e testar hipóteses, são processos de raciocínio e de inferência sistemática, que ambos os processos podem ser representados esquematicamente e que é possível distinguir caminhos melhores e piores para novos conhecimentos.

Essa linha de argumento tem muito em comum com a lógica da descoberta descrita na seção 4, acima, mas agora é explicitamente colocada contra a distinção disciplinar ligada à distinção de contexto. Existem duas maneiras principais de desenvolver esse argumento. A primeira, é conceber a descoberta em termos de raciocínio abdutivo (seção 6.1). A segunda, é conceber a descoberta em termos de algoritmos de resolução de problemas, nos quais as regras heurísticas auxiliam no processamento de dados disponíveis e aumentam o sucesso na busca de resoluções para problemas (seção 6.2). Ambas as linhas de argumentação se baseiam na concepção ampla de lógica, segundo qual a "lógica" da descoberta equivale a uma consideração esquemática dos processos de raciocínio envolvidos na geração de conhecimento.

6.1 Descoberta como abdução

Um argumento, elaborado proeminentemente por Norwood R. Hanson, é de que o ato da descoberta, aqui, o ato de sugerir uma nova hipótese, segue um padrão lógico distinto, que difere da lógica indutiva e da lógica do raciocínio hipotético-dedutivo. A lógica especial da descoberta é a lógica das inferências abdutivas ou

"retrodutivas" (*retroductive*) (HANSON, 1958). O argumento segundo o qual hipóteses científicas plausíveis e promissoras são planejadas por meio de um ato de inferências abdutiva remonta a C. S. Peirce. Essa versão da lógica da descoberta caracteriza os processos de raciocínio que acontecem *antes* de uma nova hipótese ser, em última análise, justificada. O modo de raciocínio abdutivo que conduz a hipóteses plausíveis é caracterizado como uma inferência, que se inicia com dados ou, mais especificamente, com fenômenos surpreendentes ou anômalos.

Nesta perspectiva, a descoberta é principalmente um processo de explicação de fenômenos anômalos ou fenômenos surpreendentes e espantosos. O raciocínio de cientistas procede abdutivamente de uma anomalia para uma hipótese explanatória à luz da qual os fenômenos não seriam mais surpreendentes ou anômalos. O resultado desse processo de raciocínio não é uma única hipótese específica, mas sim o delineamento de um tipo de hipótese que merece mais atenção (HANSON, 1965, p. 64). Segundo Hanson, o argumento abdutivo tem a seguinte forma esquemática (HANSON, 1960, p. 104):

1 Alguns fenômenos surpreendentes e espantosos $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$ ... são encontrados.
2 Mas $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$ ... não seriam surpreendentes se uma hipótese do tipo **H** fosse obtida. Eles se seguiriam de forma natural a partir de algo como [a hipótese] **H** e seriam explicados por ela.
3 Portanto, há boas razões para elaborar uma hipótese do tipo **H**, para sugeri-la como uma possível hipótese cuja suposições $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$ ... possam ser explicadas.

Com base em registro histórico, Hanson argumenta que várias descobertas importantes foram feitas com base no raciocínio abdutivo, como a descoberta de Kepler sobre a órbita elíptica de Marte (HANSON, 1958). Agora é amplamente aceito, no entanto, que a reconstrução de Hanson sobre esse episódio não é uma descrição historicamente adequada da descoberta de Kepler (LUGG, 1985). Mais importante, embora haja um consenso geral de que as inferências abdutivas são frequentes no raciocínio cotidiano e no científico, essas inferências não são mais consideradas como inferências **lógicas**. Mesmo se alguém aceitar a representação esquemática de Hanson do processo de identificação de hipóteses plausíveis, esse processo é "lógico" apenas no sentido mais amplo, pelo qual o termo "lógico" é entendido como sinônimo de "racional". Notavelmente, alguns filósofos chegaram a questionar a racionalidade das inferências abdutivas (KOEHLER, 1991; BREM; RIPS, 2000).

Outro argumento contra o esquema acima afirma que ele é muito permissivo. Pode haver muitas hipóteses que são explicações para os fenômenos $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$ ..., portanto, o fato de uma hipótese particular explicar os fenômenos (HARMAN, 1965; BLACKWELL, 1969), não é um critério decisivo para o seu desenvolvimento. São necessários critérios adicionais para avaliar as hipóteses geradas por inferências abdutivas.

Finalmente, vale a pena notar que o esquema do raciocínio abdutivo não explica o próprio ato de conceber uma hipótese ou um tipo de hipótese. Os processos pelos quais uma nova ideia é primeiramente articulada permanecem não analisados no esquema acima. O esquema enfatiza os processos de raciocínio pelos quais uma hipótese explanatória é avaliada em termos de seus méritos e promessas (LAUDAN, 1980; SCHAFFNER, 1993).

Em trabalhos mais recentes sobre abdução e descoberta, algumas vezes, duas noções de abdução são distinguidas: a noção comum de abdução como inferência para melhor explicação (abdução seletiva) e abdução criativa (MAGNANI, 2000, 2009). A abdução seletiva, a inferência para melhor explicação, envolve a seleção de uma hipótese a partir de um conjunto de hipóteses conhecidas. O diagnóstico médico exemplifica esse tipo de abdução. Abdução criativa, ao contrário, envolve a geração de uma hipótese nova e plausível. Isso acontece, por exemplo, em pesquisas médicas, quando a noção de uma nova doença é articulada. No entanto, ainda é uma questão em aberto se essa distinção pode ser feita, ou se há uma transição mais gradual da seleção de uma hipótese explicativa de um domínio familiar (abdução seletiva), para a seleção de uma hipótese ligeiramente modificada do conjunto familiar e a identificação de uma suposição modificada ou mais drasticamente alterada.

Outra sugestão recente é ampliar a perspectiva original de abdução de Peirce e incluir não apenas informações verbais, mas também representações mentais não verbais, como representações visuais, auditivas ou motoras. Na abordagem de Thagard, representações são caracterizadas como padrões de atividade em populações mentais (*vide* seção 9.3). A vantagem de uma perspectiva neural do raciocínio humano é a de que ela abrange características, como a surpresa, que acompanham a geração de novos insights ou representações visuais e auditivas que contribuem para isso. A surpresa, por exemplo, pode ser caracterizada como resultado de rápidas mudanças na ativação do nó em uma rede neural, representando o elemento "surpreendente" (THAGARD; STEWART, 2011). Se todas as representações mentais puderem ser caracterizadas como padrões de disparo em populações neurais, a abdução poderá ser analisada como a combinação ou a

"convolução" (*convolution*) de padrões de atividade neural a partir de padrões de atividade disjuntos ou sobrepostos (THAGARD, 2010).

6.2 Programação heurística

O interesse na lógica da descoberta motivou também a pesquisa em inteligência artificial na interseção da filosofia da ciência com a ciência cognitiva. Nessa abordagem, descoberta científica é tratada como uma forma de atividade de resolução de problemas (SIMON, 1973; NEWELL; SIMON, 1971), segundo a qual os aspectos sistemáticos da resolução de problemas são estudados dentro de uma estrutura de processamento de informação. O objetivo é esclarecer, com a ajuda de ferramentas computacionais, a natureza dos métodos utilizados para descobrir hipóteses científicas. Essas hipóteses são consideradas soluções para problemas. Filósofos que trabalham nessa tradição constroem programas de computador empregando métodos de busca heurística seletiva (LANGLEY *et al*., 1987). Na heurística computacional, programas de busca podem ser descritos como busca de soluções no chamado "espaço do problema" em um determinado domínio. O espaço do problema compreende todas as possíveis configurações nesse domínio (por exemplo, para problemas de xadrez, todos os possíveis arranjos das peças em um tabuleiro). Cada configuração é um "estado" do espaço do problema. Existem dois estados especiais, o chamado estado final, isto é, o estado a ser alcançado, e o estado inicial, isto é, a configuração no ponto de partida no qual a pesquisa começa. Existem operadores que determinam os movimentos que geram novos estados a partir do estado atual. Existem restrições de caminho que limitam os movimentos permitidos. A resolução de problemas é o processo de buscar uma solução para o problema de como alcançar o estado final a partir de um estado inicial. A princípio, todos os estados podem ser gerados a partir do emprego de operadores ao estado inicial, e então ao estado resultante, até que o estado final seja alcançado (LANGLEY *et al*., 1987, cap. 9). Uma resolução de problema é uma sequência de operações que levam do estado inicial ao estado final.

A ideia básica por trás das heurísticas computacionais é de que essas regras, que servem como diretrizes para encontrar uma solução para um determinado problema de maneira rápida e eficiente, podem ser identificadas, evitando estados indesejados no espaço do problema. Essas regras são melhores descritas como regras gerais. O objetivo de desenvolver uma lógica da descoberta, torna-se assim,

o objetivo de desenvolver uma heurística para a busca eficiente de soluções para problemas. O termo "busca heurística" indica que, ao contrário de algoritmos, os procedimentos de resolução de problemas levam à resultados meramente provisórios e plausíveis. Uma solução não é garantida, mas buscas heurísticas são vantajosas porque são mais eficientes que exaustivas buscas aleatórias de tentativa e erro. Na medida em que é possível avaliar se um conjunto de heurística é melhor, mais eficaz, que outro, a lógica da descoberta se transforma em uma teoria normativa da descoberta.

Pode-se argumentar que, por ser possível reconstruir processos importantes da descoberta científica a partir de conjuntos de heurísticas computacionais, o processo de descoberta científica pode ser considerado como um caso especial do mecanismo geral de processamento de informação. Nesse contexto, o termo "lógica" não é usado no sentido estrito de um conjunto de regras formais geralmente aplicáveis para extrair inferências, mas novamente, em um sentido amplo como um rótulo para um conjunto de regras de procedimento.

Os programas de computador que incorporam princípios de busca heurística na investigação científica simulam os caminhos que os cientistas seguiram quando eles procuravam por novas hipóteses teóricas. Programas de computador como BACON (SIMON *et al*., 1981) e KEKADA (KULKARNI; SIMON, 1988) utilizam conjuntos de heurísticas de resolução de problemas para detectar regularidades em determinados conjuntos de dados. O programa poderia discriminar, por exemplo, que os valores de um termo dependente são constantes, ou que um conjunto de valores para um termo **x** e o conjunto de valores para um termo **y** são linearmente relacionados. Portanto, o programa "inferiria" que o termo dependente sempre tem esse valor ou que existe uma relação linear entre **x** e **y**. Esses programas podem "fazer descobertas" no sentido de que podem simular descobertas bem sucedidas, como a terceira lei de Kepler (BACON) ou o ciclo de Krebs (KEKADA).

Teorias de descobertas científicas baseadas na IA ajudaram a identificar e esclarecer várias estratégias de resolução de problemas. Um exemplo dessa estratégia é a análise heurística de meios-fins, que envolve a identificação de diferenças específicas entre as situações presentes e a meta e busca por operadores (processos que mudarão a situação), que são associados às diferenças detectadas. Outra heurística importante é dividir o problema em subproblemas e começar a resolver aquele com o menor número de incógnitas a serem determinadas (SIMON, 1977). As abordagens baseadas em IA também destacaram em que medida a geração de novos conhecimentos se baseia no conhecimento existente que restringe

o desenvolvimento de novas hipóteses.

Como abordagens de descobertas científicas, as heurísticas computacionais têm algumas limitações. Mais importante, as simulações abrangem apenas certos aspectos de descobertas científicas, pois programas de computador exigem dados de experimentos reais. Esses programas não projetam novos experimentos, instrumentos ou métodos. Além disso, comparado ao espaço de problemas, especificados em heurísticas computacionais, o complexo espaço de problemas para problemas científicos são frequentemente mal definidos, e o espaço de busca relevante e o estado final, devem ser delineados antes que as suposições heurísticas possam ser formuladas (BECHTEL; RICHARDSON, 1993, cap. 1).

Os primeiros críticos de teorias de descobertas científicas baseadas na IA argumentaram que um computador não pode conceber conceitos novos, mas está confinado aos conceitos incluídos em uma dada linguagem computacional (HEMPEL, 1985, p. 119-120). Obras subsequentes mostraram que métodos computacionais podem ser usados para gerar novos resultados, levando à publicações científicas de referência na astronomia, na pesquisa sobre câncer, na ecologia e em outros campos (LANGLEY, 2000). No entanto, a pesquisa computacional mais recente sobre descoberta científica não é mais dirigida por interesses filosóficos na descoberta científica. Em vez disso, a principal motivação é contribuir com ferramentas computacionais para auxiliar os cientistas em suas pesquisas (ADDIS *et al*., 2016)

7. Anomalias e a estrutura da descoberta

Muitos filósofos sustentam que a descoberta é um assunto legítimo para a filosofia da ciência, enquanto abandonam a noção de que existe uma **lógica** da descoberta. Uma abordagem muito influente é a análise de Thomas Kuhn sobre o surgimento de novos fatos e teorias (KUHN, 1970, cap. 6). Kuhn identifica um padrão comum de descoberta como parte de sua perspectiva de mudanças científicas. Uma descoberta não é um simples ato, mas um processo extenso e complexo, que culmina em mudanças de paradigma. Paradigmas são generalizações simbólicas, compromissos metafísicos, valores e exemplares que são compartilhados por uma comunidade de cientistas, que orientam a pesquisa dessa comunidade. A ciência normal, baseada em paradigmas, não visa à novidade, mas o desenvolvimento, a extensão e a articulação de paradigmas aceitos. Uma descoberta começa com uma anomalia, isto é, com o reconhecimento de que estão sendo violadas as expectativas

induzidas por um paradigma estabelecido. O processo da descoberta envolve vários aspectos: observações de um fenômeno anômalo, tentativas de conceituá-lo e mudanças no paradigma para que a anomalia possa ser acomodada.

A marca do sucesso da ciência normal é a de que ela não faz descobertas transformadoras e, no entanto, tais descobertas ocorrem como uma consequência da ciência normal guiada por paradigmas. Quanto mais detalhado e mais desenvolvido um paradigma for, mais precisas serão suas previsões. Quanto mais precisa a formulação das expectativas dos pesquisadores, mais aptos eles estarão em reconhecer resultados anômalos e violações de expectativas:

> a novidade normalmente surge apenas para o homem que, sabendo com precisão o que deve esperar, é capaz de reconhecer que algo deu errado. Anomalia aparece apenas no contexto fornecido pelo paradigma (KUHN, 1970, p. 65).

Com base em vários exemplos históricos, Kuhn argumenta que geralmente é impossível identificar o momento exato em que algo foi descoberto ou até mesmo o indivíduo que fez a descoberta. Kuhn ilustra esses pontos com a descoberta do oxigênio (KUHN, 1970, p. 53-56). O oxigênio não havia sido descoberto antes de 1774, havia sido descoberto em 1777. Mesmo antes de 1774, Lavoisier havia notado que algo estava errado com a teoria do flogisto, mas ele foi incapaz de seguir em frente. Dois outros pesquisadores, C. W. Scheele e Joseph Priestley, identificaram, independentemente, um gás obtido pelo aquecimento de substâncias sólidas. Porém, o trabalho de Scheele permaneceu não publicado até 1777, e Priestley não identificou sua substância como um novo tipo de gás. Em 1777, Lavoisier apresentou a teoria da combustão do oxigênio, que deu origem à fundamental reconceitualização da química. Entretanto, de acordo com essa teoria, como Lavoisier a apresentou pela primeira vez, o oxigênio não era um elemento químico. Era um "princípio atômico da acidez" e o gás oxigênio era uma combinação desse princípio com o calorífico. De acordo com Kuhn, todos esses desenvolvimentos fizeram parte da descoberta do oxigênio, mas nenhum deles pode ser apontado como "o" ato da descoberta.

Em períodos pré paradigmáticos, ou em tempos de crise de paradigma, descobertas induzidas pela teoria podem acontecer. Nesses períodos, cientistas especulam e desenvolvem teorias plausíveis que podem levar a novas expectativas, experimentos e observações, com o objetivo de testar se essas expectativas podem

ser confirmadas. Embora nenhuma previsão exata possa ser feita, fenômenos que são descobertos dessa forma não são, muitas vezes, exatamente os que se esperava. Nessas situações, a exploração simultânea de novos fenômenos conjuntamente com a articulação de hipóteses plausíveis, produzem as descobertas.

Em casos como a descoberta do oxigênio, ao contrário, que ocorreu enquanto um paradigma já estava em vigor, o inesperado se torna aparente apenas de maneira lenta, com dificuldade e contra alguma resistência. As anomalias tornam-se visíveis como tais apenas de modo gradual. Leva um tempo para os investigadores reconhecerem "tanto o que algo é e o que ele é" (KUHN, 1970, p. 55). Eventualmente, um novo paradigma se estabelece e os fenômenos anômalos tornam-se os fenômenos esperados.

Estudos recentes na neurociência cognitiva sobre a atividade cerebral durante os períodos de mudança conceitual apoiam a ideia de Kuhn, segundo a qual mudança conceitual é difícil de alcançar. Esses estudos examinam os processos neurais que estão envolvidos no reconhecimento de anomalias, comparando-os com a atividade cerebral envolvida no processamento de informação compatíveis com as teorias preferidas. Os estudos sugerem que os dois tipos de dados são processados de maneira diferente (DUNBAR *et al*., 2007).

8. Metodologias de descoberta

Os defensores da perspectiva segundo a qual metodologias de descoberta existem, usam o termo "lógica" no sentido estrito de um procedimento algorítmico que gera novas ideias. Mas, do mesmo modo que teorias da descoberta científica baseadas na IA, descritas na seção 6, metodologias de descoberta científica interpretam o conceito "descoberta" como um rótulo para um processo estendido de geração e articulação de novas ideias, geralmente descrevendo esse processo em termos de resolução de problemas. Nessas abordagens, é desafiada a distinção entre contextos de descoberta e de justificação, pois entende-se que a metodologia de descoberta desempenha um papel justificativo. Os defensores de uma metodologia de descoberta geralmente baseiam-se em uma distinção entre diferentes procedimentos de justificações, justificação envolvida no processo de geração de novos conhecimentos e justificação envolvida em testá-los. Justificações consequenciais, ou "fortes", são métodos de teste. Em contraste, a justificação envolvida na descoberta é concebida como justificação gerativa (em oposição à consequencial) (seção 8.1) ou como fraca (em oposição à forte) (seção 8.2). Novamente, existe alguma ambiguidade

terminológica pois, de acordo com alguns filósofos, existem três contextos, não dois: apenas a concepção inicial de uma nova ideia (o ato criativo é o próprio contexto de descoberta, e entre ele e a justificação existe um contexto de busca separado (LAUDAN, 1980). Mas muitos defensores de metodologias de descoberta consideram o contexto de busca como parte integrante do processo da justificação. Eles mantêm a noção de dois contextos e redesenham os limites entre os contextos de descoberta e de justificação, como foram esboçados no início do século XX.

8.1 Descobertabilidade (*discoverability*)

A metodologia de descoberta tem sido caracterizada às vezes como uma forma de justificativa complementar à metodologia de teste (NICKLES, 1984, 1985, 1989). De acordo com a metodologia de teste, o suporte empírico para uma teoria resulta do teste bem sucedido das consequências preditivas derivadas dessa teoria (e das suposições auxiliares apropriadas). À luz dessa metodologia, justificação para uma teoria é "justificação consequencial", a noção de que uma hipótese é estabelecida se **novas** previsões bem sucedidas forem derivadas da teoria ou afirmação. Justificação gerativa complementa a justificação consequencial. Defensores da justificação gerativa asseguram que existe uma forma importante de justificação na ciência, que envolve o raciocínio *para* uma afirmação a partir de dados ou, de modo mais geral, a partir de resultados previamente estabelecidos.

Um exemplo clássico de uma metodologia gerativa é o conjunto de regras de Newton para o estudo da filosofia natural. De acordo com essas regras, proposições gerais são estabelecidas ao deduzi-las a partir de fenômenos. A noção de justificação gerativa procura preservar a intuição por trás de concepções clássicas da justificação por dedução. Justificação gerativa equivale à reconstrução racional do caminho da descoberta, a fim de estabelecer a sua **descobertabilidade**, caso os pesquisadores soubessem o que é conhecido agora, independentemente de como foi pensado pela primeira vez (NICKLES, 1985, 1989). A reconstrução demonstra, em retrospectiva, que a afirmação poderia ter sido descoberta dessa maneira, se as técnicas e as informações necessárias fossem disponíveis. Em outras palavras, a justificação gerativa, justificação como "descobertabilidade" ou como "descoberta potencial", justifica uma afirmação de conhecimento derivando-a de resultados que já foram estabelecidos. Embora a justificação gerativa não reescreva exatamente os passos do processo real de descoberta que foram realmente adotados, ela é uma representação

melhor das práticas reais de cientistas do que a justificação consequencial, pois cientistas tendem a formular novas afirmações a partir do conhecimento disponível. A justificação gerativa é uma versão mais fraca do ideal tradicional da justificação por dedução dos fenômenos. Justificação por dedução dos fenômenos é considerada completa se uma teoria ou afirmação for completamente determinada pelo que já se sabe. A demonstração da descobertabilidade resulta da derivação bem sucedida de uma afirmação ou teoria a partir das mais básicas e sólidas informações empíricas estabelecidas.

8.2 Avaliação preliminar

A descobertabilidade, como descrita nos parágrafos anteriores, é um modo de justificação. Assim como o teste de novas predições derivadas de uma hipótese, a justificação gerativa começa quando a fase de encontrar e articular uma hipótese digna de avaliação está chegando ao fim. Outras abordagens à metodologia da descoberta estão relacionadas diretamente com os procedimentos envolvidos na criação de novas hipóteses. O argumento a favor desse tipo de metodologia é o de que os procedimentos de elaboração de novas hipóteses já incluem elementos de avaliação. Essas avaliações preliminares foram denominadas como procedimentos "fracos" de avaliação (SCHAFFNER, 1993). Avaliações fracas são relevantes durante o processo de elaboração de uma nova hipótese. Elas fornecem razões para aceitar uma hipótese como promissora e digna de mais atenção. Avaliações fortes, por outro lado, fornecem razões para aceitar uma hipótese como (aproximadamente) verdadeira ou confirmada. Ambos os testes "gerativo" e "consequencial", como discutido na seção anterior, são procedimentos fortes de avaliação. Procedimentos fortes de avaliação são rigorosos e organizados sistematicamente de acordo com os princípios de derivação de hipóteses, ou teste de Derivação-Hipótese (D-H). Uma metodologia de avaliação preliminar, em contraste, articula critérios para a avaliação de uma hipótese antes de rigorosas derivações ou testes. Ela auxilia a decisão de levar essa hipótese a sério o suficiente para desenvolvê-la ainda mais e testá-la. Para os defensores desta versão da metodologia da descoberta, é tarefa da filosofia da ciência caracterizar conjuntos de restrições e regras metodológicas que guiam o processo complexo, anterior ao teste, de avaliação de hipóteses.

Em contraste com as abordagens computacionais discutidas acima, as estratégias de avaliação preliminares não são consideradas neutras quanto à área de conhecimento, mas específicas de campos de estudo particulares. Como a

análise de critérios para avaliação de hipóteses foi feita principalmente em relação ao estudo de mecanismo biológico, os critérios e restrições que foram propostos são aqueles que desempenham um papel na descoberta de mecanismos biológicos. Os mecanismos biológicos são entidades e atividades organizadas de modo a produzir mudanças regulares a partir de condições iniciais às condições finais (MACHAMER *et al*., 2000).

Filósofos da biologia desenvolveram um arcabouço conceitual refinado para explicar a geração e a avaliação preliminar desses mecanismos (DARDEN, 2002; CRAVER, 2002; BECHTEL; RICHARDSON, 1993; CRAVER; DARDEN, 2013). Alguns filósofos até sugeriram que a fase da avaliação preliminar fosse dividida em duas, a fase da avaliação e a fase da revisão. Segundo Lindley Darden, as fases de geração, avaliação e revisão de descrições de mecanismos podem ser caracterizadas como processos de raciocínio guiados por estratégias de raciocínio. Diferentes estratégias de raciocínio guiam as diferentes fases (DARDEN, 1991, 2002; CRAVER, 2002; DARDEN, 2009). A geração de hipóteses sobre mecanismos, por exemplo, é governada pela estratégia da "instanciação de esquema" (DARDEN, 2002). A descoberta do mecanismo de síntese proteica envolveu a instanciação de um esquema abstrato para reações químicas: reagente_1 + reagente_2 = produto. O mecanismo vigente da síntese de proteínas foi encontrado através da especificação e modificação deste esquema.

É importante avaliar o status dessas estratégias de raciocínio. Elas não são necessariamente estratégias que foram realmente usadas. Nenhuma dessas estratégias é considerada necessária para a descoberta e não são prescrições para pesquisas em biologia. Antes, essas estratégias são consideradas suficientes para a descoberta de mecanismos; elas "poderiam ter sido usadas" para chegar à descrição desse mecanismo (DARDEN, 2002). A metodologia da descoberta de mecanismos é uma extrapolação de episódios anteriores de pesquisa sobre mecanismos e o resultado de uma síntese de reconstruções racionais de vários desses episódios históricos. A metodologia da descoberta é somente pouco normativa no sentido de que as estratégias para a descoberta de mecanismos que foram identificados, até agora, podem ser úteis em futuras pesquisas biológicas. Além disso, os conjuntos de estratégias de raciocínio que foram propostos são altamente específicos. Se a análise de estratégias para a descoberta de mecanismos biológicos pode iluminar a eficiência da resolução de problemas científicos de modo mais geral, ainda é uma questão em aberto (WEBER, 2005, cap. 3).

9. Criatividade, analogia e modelos mentais

As abordagens sobre descoberta científica, apresentadas nas seções anteriores, focam na adoção, articulação e avaliação preliminar de ideias ou hipóteses antes de testes rigorosos. Eles não esclarecem como uma nova hipótese ou ideia é gerada em primeiro lugar. Mesmo entre os filósofos da descoberta, a perspectiva predominante há muito tempo é a de que o passo inicial da descoberta é melhor descrito como um momento eureka, um misterioso salto intuitivo da mente humana que não pode ser melhor analisado (STOKES, 2011).

O conceito de descoberta como formação de hipóteses, como é encapsulado na distinção tradicional entre contextos de descoberta e de justificação, não explica como são formadas novas ideias. De acordo com descrições de descobertas informadas pela biologia evolutiva, a geração de novas ideias é semelhante à variações aleatórias e cegas de processos de pensamento, que precisam ser inspecionadas pela mente crítica e avaliadas como neutras, produtivas ou inúteis (CAMPBELL, 1960; HULL, 1988). Embora a abordagem evolutiva da descoberta ofereça uma análise mais substancial sobre a descoberta científica, os principais processos pelos quais ideias aleatórias são geradas ainda são deixados sem análise.

Hoje, muitos filósofos defendem a ideia de que a criatividade não é misteriosa e pode ser submetida a análise. A psicóloga Margaret Boden oferece análises interessantes sobre o conceito de criatividade. Segundo Boden, uma nova ideia é criativa se ela for original, surpreendente e importante. Ela distingue entre criatividade psicológica (criatividade-P) e criatividade histórica (criatividade-H). Criatividade-P é o desenvolvimento de uma ideia original, surpreendente e importante para a pessoa em particular que o gerou. A criatividade-H, por outro lado, é o desenvolvimento de uma ideia radicalmente original, surpreendente e importante, isto é, gerada pela primeira vez (BODEN, 2004).

A maioria dos estudos filosóficos recentes da descoberta científica se concentra no ato de geração de novos conhecimentos. A característica distintiva desses estudos é de que eles integram abordagens da ciência cognitiva, psicologia e neurociência computacional (THAGARD, 2012; PASQUALE; POIRIER, 2016). Pesquisas recentes sobre a criatividade oferecem análises substantivas das precondições sociais e psicológicas e dos mecanismos cognitivos envolvidos na geração de novas ideias. Algumas dessas pesquisas têm como objetivo caracterizar aquelas características que são comuns a todos os processos criativos. Outros estudos têm como objetivo identificar as características que são distintivas da

criatividade científica (como oposição a outras formas de criatividade, como criatividade artística ou invenção tecnológica criativa). Estudos têm se concentrado na análise dos traços de personalidade que conduzem ao pensamento criativo e nos fatores sociais e ambientais favoráveis à descoberta (seção 9.1). Dois elementos-chave dos processos cognitivos envolvidos no pensamento criativo são analogias (seção 9.2) e modelos mentais (seção 9.3).

9.1 Condições psicológicas e sociais da criatividade

Estudos psicológicos sobre disposições comportamentais de indivíduos criativos sugerem que cientistas criativos compartilham certos traços de personalidade, incluindo confiança, abertura, prepotência, independência e introversão, bem como arrogância e hostilidade. Para um exame geral dos estudos recentes sobre traços de personalidade de cientistas criativos, consulte Feist (1999, 2006, cap. 5). A situação social também tem sido explorada como um importante recurso para a criatividade. Nessa perspectiva, as estruturas e práticas socioculturais, nas quais os indivíduos estão inseridos, são consideradas cruciais para a geração de ideias criativas. Ambas as abordagens sugerem que indivíduos criativos geralmente têm o status de estranhos (*outsiders*), eles são socialmente divergentes e se diferenciam do convencional.

O status de estranho também é uma característica de ponto de vista importante. De acordo com os teóricos do ponto de vista, pessoas com pontos de vista são politicamente conscientes e politicamente engajadas, pessoas fora do convencional. Alguns teóricos do ponto de vista sugerem explorar essa similaridade na pesquisa sobre criatividade. Como as pessoas com “ponto de vista” têm diferentes experiências e mais acesso a diferentes domínios de especialização que a maioria dos membros de uma cultura, elas podem explorar recursos conceituais valiosos para o pensamento criativo. Desse modo, a teoria do ponto de vista pode ser um recurso importante para o desenvolvimento de abordagens sociais e ambientais no estudo da criatividade (SOLOMON, 2007).

9.2 Analogia

Muitos filósofos da ciência destacam o papel da analogia no desenvolvimento de novos conhecimentos, no qual a analogia é entendida como um processo de trazer ideias que são bem entendidas em um domínio para sustentar um novo domínio (THAGARD, 1984; HOLYOAK; THAGARD, 1996). Uma importante fonte para o pensamento filosófico sobre analogia, é a concepção de modelos e analogias na elaboração e desenvolvimento de teoria de Mary Hesse. Nessa abordagem, analogias são semelhanças entre diferentes domínios. Hesse introduz a distinção entre analogias positivas, negativas e neutras (HESSE, 1966, p. 8). Se considerarmos a relação entre moléculas de gás e um modelo de gás, isto é, uma coleção de bolas de bilhar em movimento aleatório, encontraremos propriedades que são comuns em ambos os domínios (analogia positiva), bem como propriedades que podem ser atribuídas somente ao modelo mas não para o domínio de destino (analogia negativa). Existe uma analogia positiva entre as moléculas de gás e uma coleção de bolas de bilhar, pois ambas se movem aleatoriamente. Existe uma analogia negativa entre os domínios, pois as bolas de bilhar são coloridas, duras e brilhantes, mas as moléculas de gás não têm essas propriedades. As propriedades mais interessantes são aquelas do modelo sobre as quais não sabemos se são analogias positivas ou negativas. Esse conjunto de propriedades constitui a analogia neutra. Essas propriedades são significativas, pois podem levar a novas insights sobre o domínio menos familiar. A partir do nosso conhecimento sobre as conhecidas bolas de bilhar, podemos derivar novas predições sobre o comportamento das moléculas de gás que, então, poderíamos testar.

Hesse oferece uma análise mais detalhada da estrutura do raciocínio analógico, através da distinção entre analogias horizontais e verticais entre domínios. As analogias horizontais entre dois domínios dizem respeito à semelhança ou similaridade entre propriedades de ambos os domínios. Se considerarmos as ondas sonoras e de luz, existem semelhanças entre elas: o som ecoa, a luz reflete; o som é alto, a luz é brilhante, ambos são detectáveis pelos nossos sentidos. Também existem relações entre as propriedades em um domínio, como a relação causal entre o som e o tom alto que ouvimos e, analogamente, entre a luz física e a luz brilhante que vemos. Essas analogias são analogias verticais. Para Hesse, analogias verticais são a chave para a construção de novas teorias.

Analogias desempenham vários papéis na ciência. Elas não apenas contribuem para a descoberta, mas também desempenham um papel no desenvolvimento

e na avaliação de teorias científicas. As discussões atuais sobre analogia e descoberta expandiram e refinaram a abordagem de Hesse de várias maneiras. Alguns filósofos desenvolveram critérios para avaliar argumentos de analogia (BARTHA, 2010). Outras obras identificam analogias altamente significativas que foram particularmente frutíferas para o avanço da ciência (HOLYOAK; THAGARD, 1996, p. 186-188; THAGARD, 1999, cap. 9). A maioria dos analistas explora os recursos dos mecanismos cognitivos, nos quais os aspectos de um domínio ou fonte conhecidos são aplicados a um domínio de destino desconhecido, a fim de entender o que é desconhecido. De acordo com a influente teoria de múltiplas restrições do raciocínio analógico, desenvolvida por Holyoak e Thagard, os processos de transferência envolvidos no raciocínio analógico (científico e outros) são guiados ou restritos por três principais maneiras: 1) pela semelhança direta entre os elementos envolvidos; 2) pelos paralelos estruturais entre domínios de origem e de destino; bem como 3) pelos propósitos dos investigadores, isto é, os motivos pelos quais a analogia é considerada. Descoberta, a formulação de uma nova hipótese, é um desses propósitos.

Investigações *in vivo* de cientistas que raciocinam em seus laboratórios, mostraram não apenas que o raciocínio analógico é um componente essencial da prática científica, mas também que a distância entre a fonte e o destino depende do objetivo pelo qual analogias são buscadas. Cientistas que tentam corrigir problemas experimentais traçam analogias entre as fontes e os destinos de domínios altamente similares. Por outro lado, cientistas que tentam formular novos modelos ou conceitos traçam analogias entre domínios menos similares. Analogias entre domínios radicalmente diferentes, no entanto, são raras (DUNBAR, 1997, 2001).

9.3 Modelos mentais

Na ciência cognitiva atual, a cognição humana é frequentemente explorada em termos de raciocínio baseado em modelos. O ponto de partida dessa abordagem é a noção de que grande parte do raciocínio humano, incluindo o raciocínio probabilístico e causal, bem como raciocínio baseado em solução de problemas, ocorre através de modelagem mental, ao invés da aplicação de critérios lógicos ou metodológicos à um conjunto de proposições (JOHNSON-LAIRD, 1983; MAGNANI *et al*., 1999; MAGNANI; NERSESSIAN, 2002). No raciocínio baseado em modelos, a mente constrói uma representação estrutural de uma situação do mundo real ou

imaginária e a manipula. Nesta perspectiva, as estruturas conceituais são vistas como modelos e a inovação conceitual é entendida como o desenvolvimento de novos modelos através de várias operações de modelagem. O raciocínio analógico, modelagem analógica, é considerado como uma das três principais formas de raciocínio baseado em modelo que parecem ser relevantes para a inovação conceitual na ciência. Além de modelagem analógica, modelagens visual e modelos de simulação, ou experimentos de pensamento, também desempenham papéis importantes (NERSESSIAN, 1992, 1999, 2009). Essas práticas de modelagens são construtivas, pois elas ajudam no desenvolvimento de novos modelos mentais. Os elementos-chave do raciocínio baseado em modelo são o convite ao conhecimento de princípios e de restrições gerativas para modelos físicos em um domínio de origem e o uso de várias formas de abstração. A inovação conceitual resulta da criação de novos conceitos, por meio de processos que abstraem e integram domínios de fonte e de destino em novos modelos (NERSESSIAN, 2009).

Alguns críticos argumentam que, apesar da grande quantidade de trabalho sobre o tema, a noção de modelo mental não é suficientemente clara. Thagard procura esclarecer o conceito de modelo mental caracterizando-o em termos de processos neurais (THAGARD, 2010). Em sua abordagem, os modelos mentais são produzidos através de padrões complexos de disparo neural, nos quais os neurônios e as interconexões entre eles são dinâmicas e variáveis. Um padrão de neurônios disparadores se configura como uma representação quando existe uma correlação causal estável entre o padrão, ou a ativação, e o que é representado. Nessa pesquisa, as questões sobre a natureza do raciocínio baseado em modelos são transformadas em questões sobre os mecanismos cerebrais que produzem representações mentais.

As seções acima mostram que o estudo da descoberta científica se tornou parte integrante da tentativa mais ampla em explorar o pensamento criativo e a criatividade de maneira mais geral. As abordagens da filosofia naturalista combinam análise conceitual dos processos de geração de conhecimento com o trabalho empírico sobre criatividade, baseando-se forte e explicitamente às pesquisas atuais em psicologia e ciência cognitiva, às observações laboratoriais *in vivo* e, mais recentemente, às técnicas de imagem cerebral (KOUNIOS; BEEMAN, 2009; THAGARD; STEWART, 2011).

Referências

ADDIS, M.; SOZOU, P. D.; GOBET, F.; LANE, P. R. Computational scientific discovery and cognitive science theories. *In:* MUELLER, V. C. (ed.). Computing and Philosophy. **Springer**, p. 83-87, 2016.

BARTHA, P. **By Parallel Reasoning**: The Construction and Evaluation of Analogical Arguments. New York: Oxford University Press, 2010.

BECHTEL, W.; RICHARDSON, R. **Discovering Complexity.** Princeton: Princeton University Press, 1993.

BENJAMIN, A. C. The Mystery of Scientific Discovery. **Philosophy of Science**, v. 1, p. 224-36, 1934.

BLACKWELL, R. J. **Discovery in the Physical Sciences.** Notre Dame: University of Notre Dame Press, 1969.

BODEN, M. A. **The Creative Mind**: Myths and Mechanisms. London: Routledge, 2004.

BRANNIGAN, A. **The Social Basis of Scientific Discoveries**. Cambridge: Cambridge University Press, 1981.

BREM, S.; RIPS, L. J. Explanation and Evidence in Informal Argument. **Cognitive Science**, v. 24, p. 573-604, 2000.

CAMPBELL, D. Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes. **Psychological Review**, v. 67, p. 380-400, 1960.

CARMICHAEL, R. D. The Logic of Discovery. **The Monist**, v. 32, p.569-608, 1922.

CARMICHAEL, R. D. **The Logic of Discovery**. Chicago: Open Court, 1930.

CRAVER, C. F. Interlevel Experiments, Multilevel Mechanisms in the Neuroscience of Memory. **Philosophy of Science Supplement**, v. 69, p. 83-97, 2002.

CRAVER, C. F.; DARDEN, L. **In Search of Mechanisms:** Discoveries across the Life Sciences. Chicago: University of Chicago Press, 2013.

CURD, M. The Logic of Discovery: An Analysis of Three Approaches. *In:* NICKLES, T. (ed.). **Scientific Discovery, Logic, and Rationality.** Dordrecht: D. Reidel, 1980, p. 201-219.

DARDEN, L. **Theory Change in Science**: Strategies from Mendelian Genetics. New York: Oxford University Press, 1991.

DARDEN, L. Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining. **Philosophy of Science**, v. 69, p. S354-S65, 2002.

DARDEN, L. Discovering Mechanisms in Molecular Biology: Finding and Fixing Incompleteness and Incorrectness. *In:* MEHEUS, J.; NICKLES, T. (ed.). **Models of Discovery and Creativity.** Dordrecht: Springer, 2009, p. 43-55.

DUCASSE, C. J. Whewell's Philosophy of Scientific Discovery II. **The Philosophical Review**, v. 60, n. 2, p. 213-234, 1951.

DUNBAR, K. How scientists think: On-line creativity and conceptual change in science. *In:* WARD, T. B.; SMITH, S. M.; VAID, J. (ed.). **Conceptual Structures and Processes**: Emergence, Discovery , and Change. Washington: American Psychological Association Press, 1997.

DUNBAR, K. The Analogical Paradox: Why Analogy is so Easy in Naturalistic Settings Yet so Difficult in Psychological Laboratories. *In:* GENTNER, D.; HOLYOAK, K. J.; KOKINOV, B. N. (ed.). **The Analogical Mind**: Perspectives from Cognitive Science. Cambridge: MIT Press, 2001.

DUNBAR, K,; FUGELSANG, J.; STEIN, C. Do Naïve Theories Ever Go Away? Using Brain and Behavior to Understand Changes in Concepts. *In:* LOVETT, M.; SHAH, P. (ed.). **Thinking with Data**: 33rd Carnegie Symposium on Cognition. Mahwah: Erlbaum, 2007, p. 193-205.

FEIST, G. J. The Influence of Personality on Artistic and Scientific Creativity. *In:* STERNBERG, R. J. (ed.). **Handbook of Creativity**. New York: Cambridge University Press, 1999, p. 273-296.

FEIST, G. J. **The psychology of science and the origins of the scientific mind.** New Haven: Yale University Press, 2006.

GUTTING, G. Science as Discovery. **Revue internationale de philosophie**, v. 131, p. 26-48, 1980.

HANSON, N. R. **Patterns of Discovery**. Cambridge: Cambridge University Press, 1958.

HANSON, N. R. Is there a Logic of Scientific Discovery? **Australasian Journal of Philosophy**, v. 38, p. 91-106, 1960.

HANSON, N. R. Notes Toward a Logic of Discovery. *In:* BERNSTEIN, R. J. (ed.). **Perspectives on Peirce. Critical Essays on Charles Sanders Peirce.** London: Yale University Press, 1965, p. 42-65.

HARMAN, G. H. The Inference to the Best Explanation. **Philosophical Review**, v. 74, 1965.

HEMPEL, C. G. Thoughts in the Limitations of Discovery by Computer. *In:* SCHAFFNER, K. (ed.). **Logic of Discovery and Diagnosis in Medicine**. Berkeley: University of California Press, 1985, p. 115-122.

HESSE, M. **Models and Analogies in Science**. Notre Dame: University of Notre Dame Press, 1966.

HOLYOAK, K. J.; THAGARD, P. **Mental Leaps**: Analogy in Creative Thought. Cambridge: MIT Press, 1996.

HOWARD, D. Lost Wanderers in the Forest of Knowledge: Some Thoughts on the Discovery-Justification Distinction. *In:* SCHICKORE, J.; STEINLE, F. (ed.). **Revisiting Discovery and Justification. Historical and Philosophical Perspectives on the Context Distinction**. Dordrecht: Springer, 2006, p. 3-22.

HOYNINGEN-HUENE, P. Context of Discovery and Context of Justification. **Studies in History and Philosophy of Science**, v. 18, p. 501-515, 1987.

HULL, D. L. **Science as Practice**: An Evolutionary Account of the Social and Conceptual Development of Science. Chicago: University of Chicago Press, 1988.

JANTZEN, B. C. Discovery without a 'logic' would be a miracle. **Synthese**, v. 193, p. 3209-3238, 2016.

JOHNSON-LAIRD, P. **Mental Models**. Cambridge: Cambridge University Press, 1983.

KOEHLER, D. J. Explanation, Imagination, and Confidence in Judgment. **Psychological Bulletin**, v. 110, p. 499-519, 1991.

KOUNIOS, J.; BEEMAN, M. The Aha! Moment : The Cognitive Neuroscience of Insight. **Current Directions in Psychological Science**, v. 18, p. 210-216, 2009.

KORDIG, C. Discovery and Justification. **Philosophy of Science**, v. 45, p. 110-117, 1978.

KUHN, T. S. **The Structure of Scientific Revolutions**. 2. ed. Chicago: The University of Chicago Press, 1970.

KULKARNI, D.; SIMON, H. A. The processes of scientific discovery: The strategy of experimentation. **Cognitive Science**, v. 12, p. 139-176, 1988.

LANGLEY, P. The Computational Support of Scientific Discovery. **International Journal of Human-Computer Studies**, v. 53, p. 393-410, 2000.

LANGLEY, P.; SIMON, H. A.; BRADSHAW, G. L.; ZYTKOW, J. M. **Scientific Discovery**: Computational Explorations of the Creative Processes. Cambridge: MIT Press, 1987.

LAUDAN, L. Why Was the Logic of Discovery Abandoned? *In:* NICKLES, T. (ed.). **Scientific Discovery**. Dordrecht: D. Reidel, 1980, p. 173-183. v. 1.

LEPLIN, J. The Bearing of Discovery on Justification. **Canadian Journal of Philosophy**, v. 17, p. 805-814, 1987.

LUGG, A. The Process of Discovery. **Philosophy of Science**, v. 52, p. 207-220, 1985.

MACHAMER, P.; DARDEN, L.; CRAVER, C. F. Thinking About Mechanisms. **Philosophy of Science**, v. 67, p. 1-25, 2000.

MAGNANI, L. **Abduction, Reason, and Science:** Processes of Discovery and Explanation. Dordrecht: Kluwer, 2000.

MAGNANI, L. Creative Abduction and Hypothesis Withdrawal. *In:* MEHEUS, J.; NICKLES, T. (ed.). **Models of Discovery and Creativity**. Dordrecht: Springer, 2009.

MAGNANI, L.; NERSESSIAN, N. J. **Model-Based Reasoning. Science, Technology, and Values.** Dordrecht: Kluwer, 2002.

MAGNANI, L.; NERSESSIAN N. J.; THAGARD P. **Model-Based Reasoning in Scientific Discovery.** Dordrecht: Kluwer, 1999.

NERSESSIAN, N. J. How do scientists think? Capturing the dynamics of conceptual change in science. *In:* GIERE, R. (ed.). **Cognitive Models of Science.** Minneapolis: University of Minnesota Press, 1992, p. 3-45.

NERSESSIAN, N. J. Model-based reasoning in conceptual change. *In:* MAGNANI, L.; NERSESSIAN, N. J.; THAGARD, P. (ed.). **Model-Based Reasoning in Scientific Discovery.** New York: Kluwer, 1999, p. 5-22.

NERSESSIAN, N. J. Conceptual Change: Creativity, Cognition, and Culture. *In:* MEHEUS, J.; NICKLES, T. (ed.). **Models of Discovery and Creativity.** Dordrecht: Springer, 2009, p. 127-166.

NEWELL, A.; SIMON, H. A. Human Problem Solving: The State of the Theory in 1970. **American Psychologist**, v. 26, p. 145-159, 1971.

NICKLES, T. Positive Science and Discoverability. **Proceedings of the Biennial Meeting of the Philosophy of Science Association**, p. 13-27, 1984.

NICKLES, T. Beyond Divorce: Current Status of the Discovery Debate. **Philosophy of Science**, v. 52, p. 177-206, 1985.

NICKLES, T. Truth or Consequences? Generative versus Consequential Justification in Science. **Proceedings of the Biennial Meeting of the Philosophy of Science Association**, 1988, p. 393-405.

PASQUALE, J. F. de; POIRIER, P. Convolution and Modal Representations in Thagard and Stewart's Neural Theory of Creativity: A Critical Analysis. **Synthese**, v. 193, p. 1535-1560, 2016.

POPPER, K. **The Logic of Scientific Discovery**. London: Routledge,2002.

REICHENBACH, H. **Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge**. Chicago: The University of Chicago Press, 1938.

RICHARDSON, A. Freedom in a Scientific Society: Reading the Context of Reichenbach's Contexts. *In:* SCHICKORE, J.; STEINLE, F. (ed.). **Revisiting Discovery and Justification. Historical and Philosophical Perspectives on the Context Distinction**. Dordrecht: Springer, 2006, p. 41-54.

SCHAFFER, S. Scientific Discoveries and the End of Natural Philosophy. **Social Studies of Science**, v. 16, p. 387-420, 1986.

SCHAFFER, S. Making Up Discovery. *In:* BODEN, M. A. (ed.). **Dimensions of Creativity**. Cambridge: MIT Press, 1994, p. 13-511.

SCHAFFNER, K. **Discovery and Explanation in Biology and Medicine**. Chicago: University of Chicago Press, 1993.

SCHICKORE, J.; STEINLE, F. **Revisiting Discovery and Justification. Historical and Philosophical Perspectives on the Context Distinction**. Dordrecht: Springer, 2006.

SCHILLER, F. C. S. Scientific Discovery and Logical Proof. *In:* SINGER, C. J. (ed.). **Studies in the History and Method of Science.** Oxford: Clarendon, 1917, 235-289,

SIMON, H. A. Does Scientific Discovery Have a Logic? **Philosophy of Science**, v. 40, p. 471-480, 1973.

SIMON, H. A. **Models of Discovery and Other Topics in the Methods of Science**. Dordrecht: D. Reidel, 1977.

SIMON, H. A.; LANGLEY, P. W.; BRADSHAW, G. L. Scientific Discovery as Problem Solving. **Synthese**, v. 47, p. 1-28, 1981.

SMITH, G. E. The Methodology of the Principia. *In:* SMITH, G. E.; COHEN, I. B. (ed.). **The Cambridge Companion to Newton.** Cambridge: Cambridge University Press, 2002, p. 138-173.

SNYDER, L. J. Discoverers Induction. **Philosophy of Science**, v. 64, p. 580-604, 1997.

SOLOMON, M. Standpoint and Creativity. **Hypatia,** p. 226-237, 2009.

STOKES, D. Minimally Creative Thought. **Metaphilosophy**, v. 42, p. 658-681, 2011.

THAGARD, P. Conceptual Combination and Scientific Discovery. **Proceedings of the Biennial Meeting of the Philosophy of Science Association**, 1984, v. 1, p. 3-12, 1984.

THAGARD, P. **How Scientists Explain Disease**. Princeton: Princeton University Press, 1999.

THAGARD, P. How Brains Make Mental Models. *In:* MAGNANI, L.; NERSESSIAN, N. J.; THAGARD, P. (ed.). **Model-Based Reasoning in Science & Technology.** Berlin: Springer, 2010, p. 447-461.

THAGARD, P. **The Cognitive Science of Science.** Cambridge: MIT Press, 2012.
THAGARD, P.; STEWART, T. C. The AHA! Experience: Creativity Through Emergent Binding in Neural Networks. **Cognitive Science**, v. 35, p. 1-33, 2011.
WEBER, M. **Philosophy of Experimental Biology.** Cambridge: Cambridge University Press, 2005.
WHEWELL, W. **The Philosophy of the Inductive Sciences.** London: Routledge/ Thoemmes, 1996. v. 2.
ZAHAR, E. Logic of Discovery or Psychology of Invention? **British Journal for the Philosophy of Science**, v. 34, p. 2, 1983.

(III) Abdução*

Autor: Igor Douven
Tradução: Mariana Vitti Rodrigues
Revisão: Tiago L. T. Oliveira

Na literatura filosófica, o termo "abdução" é usado em dois sentidos relacionados, mas diferentes. Em ambos os sentidos, o termo refere-se a alguma forma de raciocínio explanatório. No primeiro sentido histórico, o termo refere-se ao lugar do raciocínio explanatório na **geração** de hipóteses, ao passo que, no sentido empregado com mais frequência na literatura moderna, refere-se ao lugar do raciocínio explanatório na justificativa de hipóteses. Neste último sentido, a abdução também é chamada de "inferência à melhor explicação".

Este verbete diz respeito exclusivamente à abdução,, no sentido moderno, embora exista um complemento sobre o conceito de abdução no sentido histórico, que teve origem na obra de Charles Sanders Peirce (*vide* Suplemento: **Peirce sobre Abdução**. *Vide*, também, o verbete sobre **descoberta científica**, em particular, a

---

* DOUVEN, I. Abduction. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Summer Edition. Stanford, CA: The Metaphysics Research Lab, 2017. Disponível em: https://plato.stanford.edu/archives/sum2017/entries/abduction/. Acesso em: 18 set. 2021.

The following is the translation of the entry on Abduction by Igor Douven, 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/sum2017/entries/abduction/. 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/abduction/. 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.

seção sobre **descoberta como abdução**.

A maioria dos filósofos concorda que o conceito de abdução, no sentido de inferência à melhor explicação, é um tipo de inferência que é frequentemente empregada, de uma forma ou de outra, tanto no raciocínio cotidiano quanto no científico. No entanto, a forma exata assim como o status normativo da abdução ainda são motivos de controvérsia. Esse verbete contrasta a abdução com outros tipos de inferência; indica seus usos relevantes, tanto dentro quanto fora da filosofia; considera várias caracterizações mais ou menos precisas sobre abdução; discute seu status normativo; e destaca possíveis conexões entre a abdução e a teoria bayesiana de confirmação.

1. Abdução: Uma noção geral

Você soube por acaso que, recentemente, Tim e Harry tiveram uma briga terrível que acabou com a amizade deles. Agora, alguém lhe diz que acabou de ver Tim e Harry correndo juntos. A melhor explicação que você pode pensar para esse ocorrido é que eles fizeram as pazes. Você conclui que eles são amigos novamente.

Uma manhã, você entra na cozinha e encontra um prato e um copo sobre a mesa, com migalhas de pão e um pouco de manteiga e, ao redor de um pote de geleia, você encontra um pacote de açúcar e uma caixa de leite vazia. Você conclui que um dos seus companheiros ou companheiras de casa acordou à noite para fazer um lanchinho e estava muito cansado para limpar a mesa. Você acha que sua conclusão é a melhor explicação para a cena que você observou. Para ter certeza do que ocorreu, você pensa que alguém poderia ter invadido sua casa e tenha tido tempo para dar uma beliscada durante a invasão, ou que um companheiro de casa possa ter arrumado as coisas na mesa sem ter feito um lanchinho, mas apenas para fazer você acreditar que alguém fez um lanchinho. Entretanto, essas hipóteses lhe parecem fornecer explicações muito mais artificiais dos dados observados do que àquela explicação que você havia inferido.

Caminhando pela praia, você vê algo na areia que parece ser uma imagem de Winston Churchill. Pode ser que, como nas páginas iniciais do livro de Hilary Putnam (1981), o que você vê seja, na verdade, o traço de uma formiga rastejando na praia. A explicação muito mais simples e, portanto (você acha) muito melhor, é que alguém intencionalmente desenhou uma imagem de Churchill na areia. E, de qualquer forma, é nisso que você acaba acreditando.

Nesses exemplos, as conclusões não se seguem logicamente das premissas. Por exemplo, não se segue logicamente que Tim e Harry sejam novamente amigos a partir das premissas que eles tiveram uma briga terrível que acabou com a amizade deles e que eles acabam de ser vistos correndo juntos; a conclusão não se segue, poderíamos supor, nem mesmo de toda a informação que você possui sobre Tim e Harry. Tampouco você possui dados estatísticos relevantes sobre amizades, discussões terríveis e corredores que possam justificar uma inferência a partir da informação que você possui sobre Tim e Harry para concluir que eles são amigos novamente, ou mesmo para concluir que, provavelmente (ou com uma certa probabilidade), eles são amigos novamente. O que leva você à conclusão, e o que, de acordo com um número considerável de filósofos, também justificaria essa conclusão, é precisamente o fato, sendo verdadeiro, de que Tim e Harry voltaram a ser amigos, melhor explicaria o fato de que eles foram vistos correndo juntos. (A condição de que uma hipótese, sendo verdadeira, seja capaz de explicar algo é pressuposta daqui em diante.) Observações semelhantes se aplicam aos outros dois exemplos. O tipo de inferência aqui descrita é chamado abdução ou, de certo modo mais comum nos dias de hoje, inferência à melhor explicação.

1.1 Dedução, indução, abdução

A abdução é normalmente considerada um dos três principais tipos de inferência, sendo a dedução e a indução os outros dois. A distinção entre dedução, por um lado, e indução e abdução, por outro lado, corresponde à distinção entre inferências necessárias e não necessárias. Nas inferências dedutivas, o que é inferido é **necessariamente** verdadeiro se as premissas das quais a conclusão é inferida são verdadeiras; isto é, a verdade das premissas **garante** a verdade da conclusão. Um tipo conhecido de exemplo é de inferências que seguem o seguinte esquema:

Todo A é B.
C é um A.
Então, C é um B.

Mas nem todas as inferências são desse tipo. Considere, por exemplo, a inferência “John é rico” a partir de “John vive em Chelsea” e “A maioria das pessoas

que vivem em Chelsea são ricas." Aqui, a verdade da primeira sentença não é garantida (mas apenas torna-se mais provável) pela verdade conjunta da segunda e da terceira sentenças. Em outras palavras, não é necessariamente o caso que, se as premissas são verdadeiras, então a conclusão também o será: é logicamente compatível com a verdade das premissas que John seja um dos membros da minoria de habitantes não ricos de Chelsea. O caso é semelhante em relação à sua inferência da conclusão de que Tim e Harry são amigos novamente com base na informação de que eles foram vistos correndo juntos. Talvez Tim e Harry sejam ex-parceiros de negócios que ainda tinham alguns assuntos financeiros para resolver e, por mais que quisessem evitar esses assuntos, eles decidiram compatibilizar isso com seus exercícios diários; esse cenário é compatível com a sólida decisão de nunca mais fazer as pazes.

É uma prática padrão agrupar as inferências não necessárias em **indutivas** e **abdutivas**. Inferências indutivas constituem uma classe um tanto heterogênea, mas para os presentes propósitos, elas podem ser caracterizadas como aquelas inferências que são baseadas puramente em dados estatísticos, como nas inferências sobre frequência de ocorrências observadas de dada característica específica de uma população. Um exemplo de tal inferência poderia ser:

> 96% dos estudantes do colégio flamenco falam holandês e francês.
> Louise é uma estudante do colégio flamenco.
> Então, Louise fala holandês e francês.

Informações estatísticas relevantes também podem, entretanto, ser fornecidas de maneira mais vaga, como na premissa, "A maioria das pessoas que vivem em Chelsea são ricas" (há muita discussão sobre se a conclusão de um argumento indutivo pode ser afirmada em termos puramente qualitativos ou se deve ser afirmada em termos quantitativos – por exemplo, de que sustenta que Louise fala holandês e francês com uma probabilidade de .96 – ou se isso, ocasionalmente, pode ser afirmado em termos qualitativos – por exemplo, se for alta o suficiente a probabilidade de que essa afirmação é verdadeira - e às vezes não. Nessa e em outras questões relacionadas à indução, ver Kyburg (1990, cap. 4). Também deve-se mencionar que Harman (1965) concebe a indução como um tipo especial de abdução. Ver também Weintraub (2013) para discussão.)

O simples fato de que uma inferência é baseada em dados estatísticos não é suficiente para classificá-la como uma inferência indutiva. Você pode ter observado muitos elefantes cinzas e nenhum elefante que não fosse cinza, e inferir a partir disso que todos os elefantes são cinzas, **porque isso forneceria a melhor explicação para o porquê de você ter observado tantos elefantes cinzas e nenhum elefante que não fosse cinza**. Isso seria uma instância da inferência abdutiva. Esse exemplo sugere que o melhor modo para distinguir entre indução e abdução é o seguinte: ambas são inferências **ampliativas**, o que significa que a conclusão vai além do que está (logicamente) contido nas premissas (é por isso que são inferências não-necessárias), mas em abdução há um apelo, implícito ou explícito, para considerações explanatórias, enquanto que em indução não há tal apelo; em indução há **apenas** um apelo às frequências observadas ou aos cálculos estatísticos (eu enfatizo o "apenas", porque em abdução também pode haver um apelo às frequências ou aos cálculos estatísticos, como mostra o exemplo do elefante).

Uma característica notável da abdução, compartilhada com a indução, mas não com a dedução, é que a abdução viola a **monotonicidade** da inferência, o que significa que é possível inferir abdutivamente certas conclusões a partir de um **subconjunto** de um conjunto *S* de premissas que não pode ser abdutivamente inferido a partir de *S* como um todo. Por exemplo, adicionando a premissa de que Tim e Harry são ex-parceiros de negócios que ainda possuem alguns assuntos financeiros para resolver às premissas de que eles recentemente tiveram uma briga terrível e de que eles foram vistos correndo juntos, possa não mais justificar sua conclusão de que eles são amigos novamente, mesmo se, suponhamos, que as duas últimas premissas, por si só, justifiquem a inferência. A razão para isso é que a melhor explicação concebida ao fato de Tim e Harry correrem juntos, à luz das premissas originais, possa não ser mais a melhor explicação uma vez que se adicione a informação de que eles são ex-parceiros comerciais com assuntos financeiros para resolver.

1.2 A ubiquidade da abdução

O tipo de inferência exemplificada nos casos descritos no início deste verbete parecerá completamente familiar para muitas pessoas. Filósofos, assim como psicólogos, tendem a concordar que a abdução é frequentemente empregada no raciocínio cotidiano. Às vezes, nossa confiança no raciocínio abdutivo é bem

óbvia e explícita. Em algumas tarefas cotidianas, entretanto, essa confiança pode ser tão rotineira e automática que facilmente passa despercebida. Um caso em questão pode ser nossa confiança no testemunho de outras pessoas, que se diz residir no raciocínio abdutivo; consulte Harman (1965), Adler (1994), Fricker (1994), e Lipton (1998), para defesa dessa afirmação. Por exemplo, de acordo com Jonathan Adler (1994, p. 274) "[...] a melhor explicação para o porquê de um informante afirmar que **P** é normalmente que [...] ele acredita nisso por razões devidamente apropriadas e [...] que ele pretende que eu também acredite nisso", é por isso que, normalmente, estamos justificados em confiar no testemunho de um dado informante. Isso pode estar correto, mesmo que, ao confiar no testemunho de uma pessoa, normalmente, alguém não pareça estar ciente de qualquer raciocínio abdutivo acontecendo na mente dessa pessoa. Pode-se fazer comentários semelhantes ao que alguns consideram ser outro papel, possivelmente ainda mais fundamental, da abdução na prática linguística, isto é, seu papel em determinar o que um falante quer dizer com um dado enunciado. Especificamente, argumenta-se que a decodificação de enunciados é uma questão de inferência à melhor explicação do porquê de alguém ter dito o que disse no contexto em que o enunciado foi pronunciado. Mais especificamente, autores que trabalham no campo da pragmática têm sugerido que ouvintes invocam a máxima Griceana de conversação para auxiliá-los a encontrar a melhor explicação do enunciado de um falante sempre que o conteúdo semântico do enunciado não é suficientemente informativo para os propósitos da conversação, ou é muito informativo, ou foge do assunto, ou é implausível, ou, de outro modo, é estranho ou inapropriado; veja, por exemplo Bach e Harnish (1979, p. 92), Dascal (1979, p. 167) e Hobbs (2004). Como nos casos em que se confia no testemunho do locutor, o requisito para o raciocínio abdutivo pareceria, normalmente, ocorrer no nível subconsciente.

O raciocínio abdutivo não é limitado a contextos cotidianos. Muito pelo contrário: filósofos da ciência argumentam que a abdução é o pilar da metodologia científica. Veja, por exemplo, Boyd (1981, 1984), Harré (1986, 1988), Lipton (1991, 2004) e Psillos (1999). De acordo com Timothy Williamson (2007), "[...] a metodologia abdutiva é a melhor que a ciência fornece", Já Ernan McMullin (1992) vai mais além ao chamar abdução "[...] a inferência que faz ciência". Para ilustrar o uso da abdução na ciência, consideramos dois exemplos.

No início do século XIX, foi descoberto que a órbita de Urano, um dos sete planetas conhecidos na época, desviou-se da trajetória prevista com base na teoria da gravitação universal de Isaac Newton e das pressuposições auxiliares de que

não havia outros planetas no sistema solar. Uma possível explicação foi, obviamente, que a teoria de Newton era falsa. Dado seu grande sucesso empírico por (então) mais de dois séculos, essa não parecia ser uma explicação muito boa. Dois astrônomos, John Couch Adams e Urbain Leverrier, sugeriram como alternativa (um independetemente do outro, mas quase simultaneamente) que havia um oitavo planeta, ainda não descoberto, no sistema solar; essa possibilidade, eles pensaram, fornecia a melhor explicação para o desvio na órbita de Urano. Não muito tempo depois, esse planeta, que agora é conhecido como "Netuno", foi descoberto.

O segundo exemplo diz respeito ao que agora é comumente considerado como sendo a descoberta do elétron pelo físico inglês Joseph John Thomson. Thomson conduziu experimentos em raios catódicos para determinar se eles seriam fluxos de partículas carregadas. Ele concluiu que eles, de fato, o são, raciocinando da seguinte forma:

> Como os raios catódicos carregam uma carga de eletricidade negativa, são desviados por uma força eletrostática como se fossem eletrificados negativamente e são acionados por uma força magnética exatamente da mesma maneira em que essa força atuaria em um corpo eletricamente negativo que se move ao longo do percurso desses raios, **não posso escapar da conclusão** de que eles são cargas de eletricidade negativa carregadas por partículas de matéria (THOMSON, *apud* ACHINSTEIN, 2001, p. 17).

A conclusão segundo a qual os raios catódicos são compostos por partículas carregadas negativamente não se segue logicamente dos resultados experimentais relatados, nem poderia Thomson extraí-la de nenhum dado estatístico relevante. No entanto, o fato de que ele "não pôde escapar da conclusão" se dá, podemos presumir com segurança, porque sua conclusão é a melhor explicação – nesse caso presumivelmente até a única explicação plausível – que ele poderia pensar a partir de seus resultados.

Muitos outros exemplos de empregos científicos da abdução têm sido discutidos na literatura; veja, por exemplo, Harré (1986, 1988) e Lipton (1991, 2004). Abdução é também considerada o tipo de raciocínio predominante em diagnósticos médicos: médicos tendem a adotar a hipótese que melhor explica os sintomas do paciente (*vide* JOSEPHSON; JOSEPHSON, 1994, p. 9-12); consulte, também,

Dragulinescu (2016, no prelo), sobre raciocínio abdutivo no contexto da medicina.

Por último, mas não menos importante, a abdução desempenha um papel central em alguns debates filosóficos relevantes. *Vide* Shalkowski (2010), sobre o lugar da abdução na metafísica, e, também, Bigelow (2010), Krzyżanowska, Wenmackers e Douven (2014) e Douven (2016a) para um possível papel da abdução na semântica de condicionais. Já Williamson (no prelo) para o emprego da abdução na Filosofia da lógica. Indiscutivelmente, no entanto, a abdução desempenha seu papel filosófico mais notável na epistemologia e na Filosofia da ciência, onde é frequentemente evocada para objeções aos chamados argumentos de subdeterminação. Geralmente, argumentos de subdeterminação partem da premissa de que um dado número de hipóteses é empiricamente equivalente, o que significa, de acordo com seus defensores, que a evidência, de fato, qualquer uma que possamos vir a possuir, seria incapaz de favorecer uma hipótese particular em detrimento de outras. A partir disso, deve-se concluir que não se pode justificar a crença em nenhuma das hipóteses em particular. Essa caracterização é rasa, mas servirá para os propósitos atuais; consulte Douven (2008) e Stanford (2009) para abordagens mais detalhadas dos argumentos de subtederminação. Um exemplo famoso desse tipo de argumento é o cartesiano do ceticismo global, segundo o qual a hipótese de que a realidade é mais ou menos do modo como a consideramos habitualmente é empiricamente equivalente a uma variedade das chamadas hipóteses céticas (como que somos enganados por um demônio do mal, ou que somos cérebros em uma cuba, conectados a um supercomputador). A esse respeito, confira Folina (2016). Argumentos semelhantes foram levantados como apoio ao antirrealismo científico, segundo o qual nunca teremos nenhuma justificativa para escolher entre rivais empiricamente equivalentes em relação ao que está por trás da parte observável da realidade (VAN FRAASSEN, 1980).

Respostas a esses argumentos, tipicamente, apontam para o fato de que a noção de equivalência empírica rejeita indevidamente considerações explanatórias, por exemplo, ao definir a noção estritamente em termos de hipóteses que possibilitam as mesmas previsões. Os autores de tais respostas defendem que, mesmo se algumas hipóteses possibilitem exatamente as mesmas previsões, uma delas ainda pode se configurar como sendo a melhor explicação para os fenômenos previstos. Então, se as considerações explanatórias cumprem um papel na determinação de quais inferências estamos autorizados a fazer – e, de acordo com os defensores da abdução, elas cumprem tal papel – então, ainda podemos justificar a crença na verdade (ou verdade provável, ou alguma que, como veremos abaixo, dependa da

versão da abdução presumida) de uma hipótese em um conjunto de hipóteses em que todas possibilitam as mesmas previsões. Inspirados em Bertrand Russell (1912, cap. 2), muitos epistemólogos evocam a noção de abdução para argumentar contra o ceticismo cartesiano; sua principal alegação é que, embora, a partir da fundamentação cartesiana, as hipóteses céticas possibilitem as mesmas predições que a hipótese segundo a qual a realidade é mais ou menos do modo como a consideramos habitualmente, as hipóteses céticas não são explicações igualmente boas àquelas que possibilitam as mesmas previsões; em particular, tem-se argumentado que as hipóteses céticas são consideravelmente menos simples do que a hipótese do "mundo comum". Veja, entre muitos outros, Harman (1973, cap. 8, 11), Goldman (1988, p. 205), Moser (1989, p. 161), e Vogel (1990, 2005); consulte Pargetter (1984), para uma resposta abdutiva ao ceticismo especificamente em relação às outras mentes. De modo semelhante, filósofos da ciência argumentam que estamos autorizados a acreditar na Teoria Especial da Relatividade em oposição à versão de Lorentz da teoria do éter. Mesmo que essas teorias possibilitem as mesmas previsões, a primeira é explanatoriamente superior a última (a maioria dos argumentos que sustentam essa afirmação se resumem na disputa de que a Teoria Especial da Relatividade é ontologicamente mais parcimoniosa que sua concorrente, que postula a existência de um éter. *Vide* Janssen (2002) para uma excelente discussão de várias razões pelas quais filósofos da ciência mostraram preferir a teoria de Einstein à de Lorentz).

2. Explicando a abdução

Caracterizações precisas sobre a noção de abdução são raras na literatura sobre abdução (Peirce propôs uma caracterização pelo menos razoavelmente precisa; mas, como explicado no suplemento deste verbete, ela não captura o que a maioria dos pesquisadores entende por abdução nos dias de hoje). Frequentemente, entende-se como um aspecto crucial que as considerações explanatórias dizem respeito à teoria da confirmação, ou que uma explicação bem-sucedida é uma indicação (não necessariamente infalível) da verdade. Entretanto, é nítido que essas formulações são, na melhor das hipóteses, *slogans*; e pouco esforço é necessário para compreender que elas podem ser convertidas, *prima facie*, em uma grande variedade de caracterizações plausíveis. Aqui, consideramos algumas dessas possíveis caracterizações, começando com uma que podemos chamar "versão

didática da abdução", que, como será visto, é claramente incompleta, e a seguir consideramos possíveis refinamentos dessa versão. O que essas versões têm em comum – o que não é surpreendente – é que todas elas são regras de inferência, exigindo premissas que incorporem considerações explanatórias e que concebam uma conclusão que expresse alguma afirmação sobre a verdade de uma hipótese. As diferenças entre as diversas caracterizações da abdução incidem sobre as premissas que são necessárias, ou sobre o que exatamente podemos inferir a partir delas (ou ambas as alternativas).

Nos manuais de epistemologia ou de filosofia da ciência, normalmente encontramos afirmações como a seguinte caracterização de abdução:

> **ABD1** Dada a evidência **E** e as explicações **candidatas** $\mathbf{H_1}$, ..., $\mathbf{H_n}$ sobre **E**, infere-se a verdade de que $\mathbf{H_i}$ é a melhor explicação para **E**.

Uma consideração frequente sobre essa regra, e que aponta para um problema em potencial, é a pressuposição das noções de explicação candidata e de melhor explicação, uma vez que nenhuma dessas noções possuem uma interpretação precisa. Enquanto alguns esperam que a noção de explicação candidata possa ser compreendida em termos puramente lógicos, ou ao menos em termos puramente formais, costuma-se dizer que a noção de melhor explicação deve apelar para as chamadas virtudes teóricas, como simplicidade, generalidade, e coerência com as teorias bem estabelecidas; a melhor explicação seria, então, a hipótese que, em comparação com as demais hipóteses, melhor expressa essas virtudes (*vide* THAGARD, 1978; MCMULLIN, 1996). O problema é que, atualmente, nenhuma dessas virtudes são particularmente bem compreendidas (GIERE, 1993, p. 232) chega a afirmar, radicalmente, que virtudes teóricas carecem de conteúdo real e que não desempenham nada além do que um papel retórico na ciência. Tendo em vista a recente literatura formal sobre simplicidade e sobre coerência – por exemplo Forster e Sober (1994), Li e Vitanyi (1997), e Sober (2015), sobre simplicidade e Bovens e Hartmann (2003) e Olsson (2005), sobre coerência – a primeira parte da afirmação [de Giere] se torna difícil de sustentar; além disso, Schupbach e Sprenger (2011) apresentam uma proposta de virtude explanatória em termos puramente probabilísticos. Evidências psicológicas colocam em dúvida algumas partes da segunda afirmação; veja, por exemplo, Lombrozo (2007), sobre o papel da simplicidade na avaliação dos indivíduos sobre virtude explanatória e Koslowski *et al*. (2008), sobre o papel da coerência junto ao conhecimento

de fundo para realização dessas avaliações.

Além disso, muitos dos que julgam que a noção de ABD1 corresponde a uma caracterização correta da abdução entendem que essa noção é muito forte. Alguns pensam que à abdução cabe garantir apenas a inferência de uma verdade **provável** à melhor explicação, outros entendem que a abdução garante apenas a inferência à verdade **aproximada** da melhor explicação, e outros, ainda, entendem que a abdução garante apenas a inferência à provável verdade aproximada.

O verdadeiro problema com a noção de ABD1 é, no entanto, mais profundo. Uma vez que a abdução é ampliativa – como explicado anteriormente – ela não se configura como uma regra sólida de inferência no sentido estrito da lógica, entretanto, ela é explicada exatamente nesses termos. Apesar disso, a abdução ainda pode ser **confiável**, pois, na maioria das vezes, ela possibilita uma conclusão verdadeira sempre que suas premissas são verdadeiras. Nesse sentido, uma clara condição necessária para a confiabilidade da ABD1 é inferir que, na maioria das vezes, quando é o caso que **H** melhor explica **E**, e **E** é verdadeiro, então, **H** também é verdadeira (ou **H** é aproximadamente verdadeira, ou provavelmente verdadeira, ou uma provável verdade aproximada). Mas essa condição não é suficiente para ABD1 ser confiável. Pois a ABD1 apenas assume, como premissa, que alguma hipótese é a melhor explicação para uma dada evidência **em comparação com outras hipóteses de um dado conjunto de hipóteses**. Então, se a regra é confiável, ela deve sustentar que, pelo menos normalmente, a melhor explicação relativa ao conjunto de hipóteses consideradas teria que ser, também, a melhor explicação em comparação com outras hipóteses que poderíamos ter concebido (mas que, por falta de tempo ou ingenuidade, ou por outra razão qualquer, não concebemos). Em outras palavras, deve-se sustentar que, ao menos normalmente, a explicação **absolutamente** melhor para uma dada evidência é encontrada entre as explicações candidatas que concebemos, pois ao contrário, a ABD1 pode nos levar a acreditar na "melhor hipótese de um conjunto defeituoso" (VAN FRAASSEN, 1989, p. 143).

Quão razoável é a suposição que este requisito extra seja geralmente cumprido? Presumivelmente, nem um pouco razoável. Para acreditar que esse não seja o caso, devemos pressupor algum tipo de privilégio, de modo que, quando consideramos possíveis explicações para os dados, estamos de alguma forma predispostos a encontrar, *inter alia*, a melhor explicação absoluta para esses dados. Afinal, dificilmente consideraríamos, ou ainda, dificilmente seria possível considerar, *todas* as explicações em potencial. Como van Fraassen (1989, p. 144) explicita, seria pouco plausível considerarmos *a priori* que temos esse tipo de privilégio.

Em resposta a esse questionamento, alguém poderia argumentar que o desafio em mostrar que a melhor explicação sempre, ou na maioria das vezes, está entre as hipóteses consideradas pode ser enfrentado sem a necessidade de assumir algum tipo de privilégio. Consulte Schupbach (2014) para uma resposta diferente e Dellsén (2017) para uma discussão sobre o assunto. Dadas as hipóteses que conseguimos conceber, sempre poderemos gerar um conjunto de hipóteses que, conjuntamente, esgotariam o espaço lógico. Suponha $\mathbf{H_1},\ldots,\mathbf{H_n}$ são as explicações candidatas que, até agora, fomos capazes de conceber. Em seguida, simplesmente defina $\mathbf{H_{n+1}}:= \neg\mathbf{H_1} \wedge\cdots\wedge \neg\mathbf{H_n}$ e adicione esta nova hipótese como mais uma explicação candidata àquelas que já tínhamos concebido. Obviamente, o conjunto $\{\mathbf{H_1},\ldots,\mathbf{H_{n+1}}\}$ é exaustivo, e um de seus elementos deve ser verdadeiro. Seguir esse simples procedimento pareceria suficiente para garantir que nunca perderíamos a explicação absolutamente melhor. Veja Lipton (1993), para uma proposta nesse sentido.

Infelizmente, há um problema. Por mais que haja muitas hipóteses $\mathbf{H_j}$ que impliquem $\mathbf{H_{n+1}}$ e, uma vez formuladas, poderiam ser avaliadas como sendo melhores explicações para os dados em relação à melhor explicação dentre as explicações candidatas com as quais começamos, $\mathbf{H_{n+1}}$, em si mesma, será dificilmente informativa; de fato, em termos gerais não fica claro quais seriam suas consequências empíricas. Por exemplo, suponha que tenhamos, como explicações rivais, a teoria especial da relatividade e a versão de Lorentz da teoria do éter. Assim, de acordo com a proposta acima, poderíamos adicionar às nossas explicações candidatas que nenhuma das duas teorias é verdadeira. Mas certamente essa hipótese adicional será classificada como possuindo pouco poder explanatório - isso se classificada, o que parece duvidoso, dado que não é totalmente claro quais seriam suas consequências empíricas. Isso para não dizer que, talvez, o procedimento sugerido nunca funcione. A questão é que, geralmente, isso fornecerá pouca garantia de que a melhor explicação está entre as explicações candidatas que consideramos.

Uma resposta mais promissora para o já mencionado “argumento do conjunto defeituoso de hipóteses” se inicia com a observação de que esse argumento tira proveito da assimetria ou incongruência peculiar da ABD1. Essa regra de inferência possibilita uma conclusão absoluta – que uma dada hipótese é verdadeira – com base em uma premissa comparativa, a saber, que aquela hipótese particular é a melhor explicação para a evidência relativa às outras hipóteses disponíveis (KUIPERS, 2000, p. 171). Essa incongruência não é evitada ao substituir “verdade” por “verdade provável” ou “verdade aproximada”. Para evitá-la, temos, em geral, duas opções.

A primeira opção é modificar a regra para que ela exija uma premissa absoluta. Por exemplo, segundo Alan Musgrave (1988) ou Peter Lipton (1993), pode-se exigir que hipótese da qual a verdade é inferida não seja apenas a melhor das explicações potenciais disponíveis, mas seja, também, a mais **satisfatória** (Musgrave) ou uma hipótese **boa o suficiente** (Lipton), formando a seguinte variação da ABD1:

> **ABD2** Dada a evidência **E**, e as explicações candidatas $\mathbf{H_1},\ldots,\mathbf{H_n}$ para **E**, infere-se a verdade de que $\mathbf{H_i}$, a qual melhor explica **E** desde que se garanta que $\mathbf{H_i}$, é uma hipótese satisfatória/boa o suficiente qua explicação.

Não é preciso dizer que a ABD2 precise ser complementada por um critério, que ainda não temos, que indique quão satisfatória sejam suas explicações, ou se são boas o suficiente.

Em segundo lugar, pode-se formular uma versão simétrica ou congruente da abdução que autorize, dada uma premissa comparativa, apenas uma conclusão comparativa; esta opção, por sua vez, pode também ser caracterizada de diferentes modos. Abaixo, têm-se uma das formas dessa caracterização, que tem sido proposta e defendida na obra de Theo Kuipers (1984, 1992, 2000).

> **ABD3** Dada a evidência **E**, e as explicações candidatas $\mathbf{H_1},\ldots,\mathbf{H_n}$ para **E**, se $\mathbf{H_i}$ explica **E** melhor do que qualquer uma das outras hipóteses, infere-se que $\mathbf{H_i}$ está mais próxima da verdade do que qualquer uma das outras hipóteses.

Claramente, ABD3 requer uma abordagem do grau de aproximação da verdade; atualmente, muitas dessas abordagens são oferecidas (NIINILUOTO, 1998).

Uma característica notável das versões congruentes da abdução aqui consideradas é que elas não se apoiam na suposição de um privilégio implausível por parte da pessoa que raciocina que, como vimos, a ABD1 implicitamente pressupõe. Outra característica é que, se alguém pode ter certeza de que, não importa quantas explicações candidatas para os dados não foram concebidas, nenhuma delas equivaleria a melhor explicação **concebida**, assim as versões congruentes permitem realizar exatamente a mesma inferência que a ABD1 (supondo que não se tenha certeza de

que nenhuma explicação potencial seja tão boa quanto a melhor explicação que se tenha concebido, mesmo se a última não for nem satisfatória nem suficientemente boa).

Como mencionado, existe um amplo consenso segundo o qual pessoas habitualmente se apoiam no raciocínio abdutivo. Em qual das regras acima mencionadas as pessoas se apoiam **exatamente**? Ou, ainda, poderiam elas se apoiam em outra regra de inferência? Ou poderiam, em alguns contextos, se apoiar em uma versão e, em outros contextos, em outra versão (DOUVEN, 2017)? Argumentações filosóficas não são capazes de responder a essas questões. Nos últimos anos, psicólogos experimentais começaram a se atentar para o papel que os humanos atribuem aos aspectos explicativos do raciocínio. Por exemplo, Tania Lombrozo e Nicholas Gwynne (2014) descrevem experimentos que mostram **como** uma propriedade de uma dada classe de objetos é explicada – se mecanicamente, pela referência a suas partes e processos, ou funcionalmente, pela referência a suas funções e propósitos – considerando o quão provável somos capazes de generalizar uma propriedade para outra classe de objetos. Confira, também, Sloman (1994) e Williams e Lombrozo (2010). Igor Douven e Jonah Schupbach (2015a,b) apresentam evidências experimentais para o efeito segundo o qual a atualização de probabilidade das pessoas tende a ser influenciada pelas considerações explanatórias que divergem de diferentes modos de atualização estritamente bayesianas. Já Douven (2016b) mostra que, nos experimentos acima mencionados, os participantes que atribuíam maior valor às considerações explicativas tendiam a ser mais precisos, conforme determinado nos termos de uma regra de pontuação padrão. Veja Lombrozo (2012, 2016) para uma noção geral sobre experimentos recentes que são relevantes para o estudo da explicação e da inferência. Mesmo assim, não há nada na literatura em psicologia, até o momento, que indique possíveis respostas às questões acima mencionadas.

Em relação à questão normativa a respeito de em qual das regras de inferência acima mencionadas **devemos** nos apoiar (se tivermos que nos apoiar em alguma forma de abdução), em que uma argumentação filosófica deveria ser capaz de nos auxiliar, a situação dificilmente é melhor. À luz do argumento do conjunto defeituoso de hipóteses, ABD1 não parece ser uma caracterização muito boa. Outros argumentos, contrários à abdução, são independentes de uma explicação precisa da regra de inferência; veremos abaixo que esses argumentos são insuficientes. Por outro lado, os argumentos a favor da abdução – alguns dos quais também serão discutidos abaixo – não diferenciam entre suas versões específicas. Então, supondo que, habitualmente, as pessoas de fato se apoiam na abdução, é preciso considerar

a questão ainda em aberto sobre em qual versão (ou versões) da abdução elas se apoiam. Igualmente, supondo que seja racional para as pessoas raciocinar por abdução, é preciso considerarmos a questão ainda em aberto sobre em qual versão, ou talvez em quais versões, da abdução elas devam, ou pelo menos tenham permissão para confiar.

3. O status da abdução

Mesmo sendo verdade que, habitualmente, nos apoiamos na abdução, poderíamos ainda ser questionados se essa prática é racional. Por exemplo, estudos experimentais mostram que quando indivíduos conseguem pensar sobre uma explicação de algum evento possível, eles tendem a superestimar a probabilidade real de ocorrência desse evento. Confira Koehler (1991), para um exame de alguns estudos como esse, e, também, Brem e Rips (2000). De forma mais notável, Lombrozo (2007) mostra que, em algumas situações, as pessoas tendem a superestimar grosseiramente a probabilidade de explicações simples em comparação com explicações mais complicadas. Embora esses estudos não sejam diretamente relacionados com a noção de abdução em nenhuma de suas versões discutidas até agora, eles sugerem, no entanto, que a análise de considerações explanatórias do ato de raciocinar nem sempre é a melhor forma de pesquisa (deve-se notar que os experimentos de Lombrozo **estão** diretamente relacionados com alguma proposta concebida com o intuito de explicar a abdução de acordo com o arcabouço conceitual bayesiano; ver Seção 4). Entretanto, as observações mais pertinentes sobre o status normativo da abdução são encontradas, até o momento, na literatura filosófica. Esta seção discute as principais críticas levantadas contra a abdução, assim como os argumentos mais significativos concebidos em sua defesa.

3.1 Críticas

Já nos deparamos com o argumento do conjunto defeituoso de hipóteses que, como vimos, é uma crítica válida para a ABD1, mas sem valor em relação a várias, como denominamos, regras congruentes da abdução. Consideramos aqui duas objeções que devem ser compreendidas de forma mais geral. A primeira objetiva desafiar a ideia central que fundamenta a abdução; a segunda não é tão

geral quando a primeira, mas ainda pretende minar uma ampla classe de explicações candidatas da abdução. Ambas as objeções são propostas por Bas van Fraassen.

A primeira objeção possui a premissa de que, como parte do significado de "explicação", se uma teoria é mais explicativa que outra, a primeira deve ser mais informativa do que a última (VAN FRAASSEN, 1983, Seção 2). O referido problema, então, é que "[...] uma questão lógica elementar indica que uma teoria mais informativa não pode ser mais provável de ser verdadeira [e, assim] tentativas de descrever uma base indutiva ou evidencial por meio de características que requerem informação (como em 'inferência à melhor explicação') devem ou contradizer-se, mutuamente ou estar equivocadas (VAN FRAASSEN, 1989, p. 192). A questão lógica elementar deve ser "[...] mais [óbvia] ... no caso paradigmático em que uma teoria é uma extensão de outra: claramente, a extensão tem mais chances de ser falsa" (VAN FRAASSEN, 1985, p. 280).

É importante considerar, entretanto, que em qualquer outro tipo de caso que não seja o caso "paradigmático", a questão supostamente elementar não é de todo óbvia. Por exemplo, é completamente obscuro o sentido em que a teoria especial da relatividade "[...] tem mais chances de ser falsa" em relação a teoria do éter de Lorentz, dado que ambas possibilitam as mesmas previsões. E, ainda, a primeira é geralmente concebida como sendo superior, *qua* explicação, em relação à última (se van Fraassen objetasse que a primeira não é realmente mais informativa que a última, ou, de qualquer forma, mais informativa em um sentido apropriado – qualquer que seja – então deveríamos certamente recusar a premissa de que, para ser mais explicativa, uma teoria deve ser mais informativa).

A segunda objeção oferecida em van Fraassen (1989, cap. 6), diz respeito às versões probabilísticas da abdução. De acordo com essa objeção, as regras devem ou corresponder à regra de Bayes, e então serem redundantes, ou estarem em desacordo com essa regra, e então, com base no argumento dinâmico do caderno de aposta holandês concebido por Lewis, como descrito em Teller (1973), serem probabilisticamente incoerentes, o que significa que é possível levar alguém a avaliar uma grande quantidade de apostas que, juntas, garantiriam uma perda financeira, seja como for; e, como van Fraassen argumenta, seria irracional seguir uma regra que possua tal característica.

No entanto, essa objeção não se sai melhor que a primeira. Pois, como Patrick Maher (1992) e Brian Skyrms (1993) indicam, um fracasso em um aspecto pode ser superado por um benefício em outro. Pode ser, por exemplo, que alguma versão probabilística da abdução tenha melhor desempenho, pelo menos em nosso

universo, do que uma regra de Bayes que, normalmente, se aproxima mais rapidamente da verdade no sentido de que é mais rápido atribuir uma alta probabilidade (entendida como probabilidade acima de um determinado limite de valor) para a hipótese verdadeira (*vide* DOUVEN, 2013; DOUVEN; WENMACKERS, no prelo; CLIMENHAGA, no prelo). Se esse for o caso, então seguir essa regra ao invés da regra bayesiana pode trazer vantagens que, talvez, não sejam prontamente expressas em termos monetários, mas que ainda sim devam ser consideradas na hora de decidir qual regra adotar. Em resumo, não é tão claro se a adoção de uma regra probabilisticamente incoerente deva ser considerada irracional.

Por outro lado, Douven (1999) argumenta que a questão a respeito da coerência de uma regra probabilística não pode ser resolvida independentemente da consideração de quais outras regras epistêmicas e decisões teóricas são concomitantemente empregadas; a coerência deve ser entendida como uma propriedade do conjunto de regras epistêmicas e de decisões teóricas, não como regras epistêmicas (como as regras probabilística de mudança de crença) isoladas. No mesmo artigo, um conjunto coerente de regras é descrito incluindo uma versão probabilística da abdução. Confira Kvanvig (1994), Harman (1997), Leplin (1997), Niiniluoto (1999) e Okasha (2000), para diferentes respostas à crítica de van Fraassen sobre as versões probabilísticas da abdução.

3.2 Defesas

Dificilmente alguém se comprometeria, atualmente, com uma concepção de verdade que postule uma conexão necessária entre força explanatória e verdade – porque isso estabeleceria, por exemplo, que a superioridade explanatória é necessária para obtenção da verdade. Como resultado, defesas a priori da abdução não são consideradas. De fato, todas as defesas concedidas até agora são de natureza empírica no sentido que apelam para os dados que supostamente baseiam a afirmação segundo a qual, de alguma forma, a abdução é uma regra sólida de inferência.

O melhor argumento deste tipo conhecido foi desenvolvido por Richard Boyd (1981, 1984, 1985), nos anos de 1980. Ele inicia seu argumento ressaltando a dependência teórica da metodologia científica, que compreende métodos de delineamento de experimentos, que analisa dados, que escolhe entre hipóteses rivais, e assim por diante. Por exemplo, ao considerar possíveis fatores ambíguos dos quais um delineamento experimental deve ser protegido, cientistas se apoiam

veementemente em teorias já aceitas. Em seguida, Boyd chama a atenção para a aparente confiabilidade dessa metodologia que, sobretudo, produziu, e continua produzindo, teorias impressionantemente precisas. Em particular, ao se basear nessa metodologia, cientistas são capazes, já há algum tempo, de encontrar teorias cada vez mais instrumentalmente adequadas. Assim, Boyd argumenta que essa confiabilidade da metodologia científica é mais bem explicada pela pressuposição de que as teorias nas quais os cientistas se baseiam são pelos menos aproximadamente verdadeiras. A partir disso e do fato de que essas teorias são, em sua maioria, geradas por raciocínio abdutivo, ele conclui que a abdução deve ser uma regra confiável de inferência.

Alguns críticos têm acusado este argumento de incorrer em circularidade. Especificamente, alega-se que este argumento se apoia na premissa – segundo a qual a metodologia científica é informada por teorias aproximadamente verdadeiras - que, por sua vez, se apoia em uma inferência para a melhor explicação para dar conta de sua plausibilidade. Sendo que a confiabilidade neste tipo de inferência é exatamente o que está em jogo (LAUDAN, 1981; FINE 1984).

A esse tipo de crítica, Stathis Psillos (1999, cap. 4) responde apelando para a distinção atribuída a Richard Braithwaite, a saber, a distinção entre circularidade de premissa e circularidade de regra. Um argumento apresenta uma circularidade de premissa se sua conclusão se encontra dentre suas premissas. Um argumento que contém uma circularidade de regra, ao contrário, é um argumento no qual a conclusão afirma algo sobre uma regra inferencial que é utilizada no mesmo argumento. Psillos insiste que o argumento concebido por Boyd contém uma circularidade de regra, mas não uma circularidade de premissa, e, argumentos que apresentam circularidade de regra, afirma Psillos, não apresentam **necessariamente** uma circularidade viciosa (mesmo se um argumento que contenha circularidade de premissa sempre apresente uma circularidade viciosa). Mais precisamente, em sua opinião, um argumento para a confiabilidade de uma dada regra **R** que, essencialmente dependa de **R** como um princípio de inferência, não é vicioso; uma vez que o uso de **R** não garante uma conclusão positiva sobre a confiabilidade de **R**. Psillos afirma que esta condição é alcançada no argumento de Boyd. Enquanto Boyd conclui que as teorias de fundo nas quais a metodologia científica se apoia são aproximadamente verdadeiras com base em uma etapa abdutiva, o uso da própria abdução não garante a verdade de sua conclusão. Afinal, admitir o emprego da abdução não faz nada para garantir que a melhor explicação sobre o sucesso da metodologia científica corresponda à verdade aproximada de teorias relevantes. Assim, conclui Psillos, o

argumento de Boyd ainda permanece.

Mesmo se o uso da abdução, no argumento de Boyd, possa levar à conclusão segundo a qual a abdução **não** é confiável, pode-se ainda duvidar sobre a possibilidade de a abdução apresentar circularidade de regra. Suponha que alguma comunidade científica tenha se baseado não na abdução, mas em uma regra que podemos chamar "Inferência para a pior explicação" (IPE), uma regra que autoriza a inferência para a **pior** explicação a partir dos dados disponíveis. Podemos assumir com segurança que essa regra, na maioria das vezes, levaria à adoção de teorias muito malsucedidas. Entretanto, a referida comunidade poderia justificar o uso da IPE por meio do seguinte raciocínio: "Teorias científicas tendem a ser extremamente malsucedidas. Essas teorias foram geradas por meio da aplicação da IPE. Que a IPE é uma regra confiável de inferência – isto é, uma regra de inferência que, na maioria das vezes, extrai conclusões verdadeiras de premissas verdadeiras – é certamente a pior explicação para o fato de que nossas teorias são muito malsucedidas. Então, ao empregarmos a IPE, podemos concluir que a IPE é uma regra confiável de inferência". Embora essa seja uma conclusão absurda, o argumento que leva a essa conclusão não pode ser acusado de apresentar uma circularidade viciosa mais do que se poderia acusar o argumento de Boyd sobre a confiabilidade da abdução (se Psillos estiver correto). Parece, então, que deve haver algo a mais de errado com a circularidade de regra.

É razoável notar que para Psillos, o fato de um argumento que apresente circularidade de regra não garantir uma conclusão positiva sobre a regra em questão, não é suficiente para que esse argumento seja válido. Outra condição necessária seria "[...] que não se deva possuir nenhum motivo para desconfiar da confiabilidade da regra, pois não há nada disponível no momento que possa fazer com que se desconfie da regra" (PSILLOS, 1999, p. 85). E há muitos motivos para desconfiar da confiabilidade da IPE; de fato, o argumento acima **supõe** que ela não é confiável. No entanto, duas questões aparecem. Primeiro, por que deveríamos aceitar essa condição adicional? Segundo, realmente não teríamos **nenhum** motivo para duvidar da confiabilidade da abdução? Certamente, **algumas** das inferências abdutivas que realizamos nos levam a aceitar **falsidades**. Quantas falsidades podemos aceitar baseados na abdução antes de podermos legitimamente começar a desconfiar dessa regra? Nenhuma resposta satisfatória foi dada a essas questões.

De qualquer forma, mesmo que a circularidade de regra não seja nem viciosa nem problemática, ainda podemos questionar como o argumento de Boyd supera a crítica da abdução, dado que ele se baseia na abdução. Psillos, entretanto,

deixa claro que o objetivo da argumentação filosófica não é sempre, e em muitos casos não precisa ser, a de convencer um oponente da posição de alguém. Às vezes, o objetivo é, de maneira mais modesta, garantir ou assegurar a si mesmo que a posição que alguém endossa, ou a tentativa de endossar tal posição, esteja correta. No caso em questão, não precisamos compreender o argumento de Boyd como uma tentativa de convencer os oponentes sobre a confiabilidade da abdução. Ao invés disso, é preciso refletir sobre maneiras de justificar a regra de acordo com a perspectiva de qualquer um que já seja simpático à abdução (PSILLOS, 1999, p. 89).

Há, também, tentativas mais diretas de argumentar a favor da abdução, a saber, por meio da indução enumerativa. A ideia em comum dessas tentativas é que todo emprego recém analisado e bem-sucedido da abdução, como a descoberta de Netuno, cuja existência foi postulada com base em explicações (ver seção 1.2), fortalece a hipótese de que a abdução é uma regra confiável de inferência, de modo que toda nova observação de um corvo preto fortalece, de alguma forma, a hipótese segundo a qual todos os corvos são pretos. Uma vez que essa inferência não envolve raciocínio abdutivo, é mais provável que esse tipo de argumento também atraia os descrentes da abdução. Consulte Harré (1986, 1988), Bird (1998, p. 160), Kitcher (2001), e Douven (2002) para sugestões nessa linha de argumentação.

4. Abdução versus Teoria Bayesiana da Confirmação

Na década passada, a teoria bayesiana da confirmação foi consolidada como uma proposta dominante sobre a confirmação; atualmente, não se pode discutir questões relacionadas à confirmação sem deixar claro se, e neste caso, o motivo pelo qual a posição considerada é diferente do pensamento bayesiano tradicional. A abdução, em qualquer de suas versões, atribui um papel de confirmação teórico para a explicação: considerações explanatórias contribuem para que algumas hipóteses se tornem mais confiáveis, e outras menos. Em contraste, a teoria bayesiana da confirmação não faz nenhuma referência ao conceito de explanação. Isso implica que a abdução difere veementemente da doutrina predominante em teoria da confirmação? Recentemente, muitos autores têm argumentado que não somente a abdução é compatível com o bayesianismo, mas se figura como um complemento muito necessário. Até o momento, a defesa mais completa disponível dessa análise foi concebida por Lipton (2004, cap. 7); em seu entendimento, bayesianos também devem ser “explicacionistas” (o nome dado por Lipton aos

defensores da abdução). Para outras defesas, confira Kasha (2000), McGrew (2003), Weisberg (2009), e Poston (2014, cap. 7) e, para discussões sobre o assunto, veja Roche e Sober (2013, 2014) e McCain e Poston (2014).

Essa questão exige alguns esclarecimentos. O que significa, para um bayesiano, ser explicacionistas? A fim de empregar a regra de Bayes e determinar a probabilidade de **H** depois da observação de **E**, o agente bayesiano terá que determinar a probabilidade de **H** condicionada a **E**. Para isso, ele precisa atribuir probabilidades não condicionadas para **H** e **E** bem como atribuir a probabilidade de **E** uma vez que se tenha **H**; as duas primeiras probabilidades não condicionadas são, na maioria das vezes, chamadas "probabilidades prévias" de, respectivamente, **H** e **E**, a última probabilidade condicionada é chamada "verossimilhança" de **H** em relação a **E** (esta é a história oficial do bayesianismo. Nem todos os simpatizantes do bayesianismo aceitam essa história. Por exemplo, de acordo com alguns pesquisadores, é mais razoável pensar que as probabilidades condicionais são fundamentais e que derivamos, com base nelas, as probabilidades não condicionadas; ver Hájek (2003), e as referências nele citadas. Como os bayesianos determinam esses valores? **Como é bem conhecido, a teoria da probabilidade fornece mais probabilidades uma vez que já temos acesso às probabilidades de alguns dados**; essa teoria não concede as probabilidades a partir do nada. É claro que, quando **H** implica **E** ou a negação de **E**, ou quando **H** é uma hipótese estatística que concede alguma probabilidade sobre **E**, então a verossimilhança se segue "analiticamente". Essa afirmação assume uma versão do Princípio Principal de Lewis (1980), como há controvérsias se esse princípio é ou não analítico, utilizamos as aspas. Mas esse não é sempre o caso, e mesmo se fosse, não se colocaria a questão de como determinar as probabilidades prévias. Aqui é onde, de acordo com Lipton, a abdução entra. Em sua proposta, os bayesianos devem determinar as probabilidades prévias e, se possível, a verossimilhança com base nas considerações explanatórias.

Como, exatamente, considerações explanatórias orientariam a escolha de probabilidades prévias? A resposta para essa questão não é tão simples como se possa em princípio imaginar. Suponha que você está considerando quais probabilidades prévias atribuir a um conjunto de hipóteses rivais e que você siga a sugestão de Lipton. Como você faria isso? Uma resposta óbvia – apesar de vaga – poderia ser: quaisquer que sejam as probabilidades prévias que você venha a atribuir, você deveria atribuir uma probabilidade mais alta para a hipótese que melhor explique os dados disponíveis em relação a quaisquer hipóteses rivais (desde que haja uma melhor explicação). Note, entretanto, que seu colega, que é bayesiano, mas acha

que confirmação não tem nada a ver com explicação, pode muito bem atribuir uma probabilidade prévia para a melhor explicação que seja ainda mais alta do que aquela que você atribuiu para sua hipótese. De fato, as probabilidades prévias que ele atribuiu para melhor explicação podem ainda ser consistentemente maiores do que as suas, não porque, na opinião dele, a explicação é de alguma forma relacionada a confirmação – ele acha que não é – mas apenas porque sim. Nesse contexto, "porque sim" se torna um motivo perfeitamente legítimo, porque qualquer motivo que se estabeleça como probabilidade prévia é um motivo legítimo para os padrões bayesianos. Segundo a epistemologia bayesiana convencional, as probabilidades prévias (e às vezes as verossimilhanças) são discutíveis, o que significa que uma atribuição de probabilidades prévias pode ser tão boa quanto outra atribuição, uma vez que ambas sejam coerentes (isto é, que obedeçam aos axiomas da teoria da probabilidade). A recomendação de Lipton para que os bayesianos se tornem explicacionistas é completamente geral. Mas o que seu colega faria de diferente se ele quisesse seguir tal recomendação? Deveria ele atribuir a mesma probabilidade prévia a qualquer melhor explicação que você, o colega explicacionistas dele, atribuiu, ou seja, diminuir as probabilidades prévias que ele havia atribuído à melhor explicação? Ou, ao invés disso, ele deveria atribuir probabilidades prévias ainda mais altas para a melhor explicação do que àquelas que ele havia atribuído?

Talvez, a proposta de Lipton não vise abordar aqueles que já atribuem as maiores probabilidades prévias às melhores explicações, mesmo que o façam baseados em argumentos que nada tenham a ver com explicação. A ideia pode ser que, na medida em que se atribua as mais altas probabilidades prévias a essas hipóteses, o procedimento terá êxito, ou pelo menos terá mais êxito do que se não as atribuíssem, não importando os motivos para a atribuição dessas probabilidades prévias. Ao que tudo indica, a solução para a questão sobre como considerações explanatórias orientam a escolha de probabilidades prévias poderia se dar a partir do comprometimento em se atribuir probabilidades prévias mais altas para a melhor explicação do que para suas rivais, se isso já não tiver sido feito. Se essa atribuição já estiver sendo feita, deverá assim continuar.

Como interpelação, deve-se notar que, de acordo com o uso padrão do bayesianismo, o termo "probabilidades prévias" não se refere, necessariamente, aos graus de crença que alguém possui antes da coleta de **qualquer** dado. Se já há algum dado, então, claramente, pode-se atribuir probabilidades prévias mais altas para hipóteses que melhor explique os dados então disponíveis. Entretanto, pode-se falar de maneira sensata em "melhores explicações" antes mesmo que

quaisquer dados sejam conhecidos. Por exemplo, uma hipótese pode ser considerada como sendo a melhor explicação em detrimento de qualquer uma de suas rivais porque ela exige menos cálculos complicados, ou porque ela é descrita apenas em termos de conceitos familiares, o que não acontece com as outras. De maneira mais geral, tais considerações podem ser baseadas no que Kosso (1992, p. 30) denomina **características internas** de hipóteses ou teorias, ou seja, características que "podem ser avaliadas sem que se observe o mundo".

Uma resposta mais interessante para a questão acima sobre como considerações explanatórias orientam a escolha de probabilidades prévias foi concebida por Jonathan Weisberg (2009). Dissemos que os bayesianos tradicionais consideram uma atribuição de probabilidades prévias como sendo tão boa quanto qualquer outra. Entretanto, os chamados bayesianos objetivos não pensam assim. Esses bayesianos entendem que as probabilidades prévias, para serem aceitáveis, devem obedecer a princípios que vão além dos axiomas da probabilidade. Os bayesianos objetivos não concordam em relação a quais princípios adicionais devem, exatamente, ser obedecidos; mas, pelo menos por algum tempo, eles concordaram que o Princípio da Indiferença é um deles. De forma geral, esse princípio indica que, na falta de razões antagônicas, temos que atribuir igual probabilidade prévia às hipóteses concorrentes. Como é bem sabido, entretanto, em sua forma original, o Princípio da Indiferença pode levar a atribuição inconsistente de probabilidades e, assim sendo, dificilmente pode ser aconselhado como princípio de racionalidade. O problema é que, tipicamente, temos muitas maneiras de particionar o espaço lógico que parecem plausíveis, considerando o problema em questão, mas nem todas dessas maneiras levam à atribuição das mesmas probabilidades prévias, mesmo assumindo o princípio da indiferença. A proposta de Weisberg corresponde à afirmação de que considerações explanatórias podem favorecer algumas das partições do espaço lógico em detrimento de outras. Talvez, nem sempre vamos acabar com uma partição única na qual o Princípio da Indiferença possa ser empregado, mas já seria algum progresso se terminássemos com apenas algum conjunto de partições. Pois assim, ainda poderíamos chegar às nossas probabilidades prévias, de maneira razoável, por meio de dois passos, a saber, primeiro atribuindo separadamente o Princípio da Indiferença para as partições do espaço lógico, obtendo, assim diferentes atribuições de probabilidades prévias e, em seguida, considerando uma média ponderada das probabilidades prévias obtidas, em que as ponderações também dependam das considerações explanatórias. O resultado seria, novamente, uma função de probabilidade – a única função de

probabilidade prévia correta, segundo Weisberg.

A proposta é intrigante do modo como foi caracterizada mas, como Weisberg admite, na sua forma atual, ela não explica muita coisa. Por um lado, não está claro como, exatamente, as considerações explanatórias determinariam os cálculos exigidos no segundo passo da proposta. Por outro lado, pode ser inútil esperar que a reflexão sobre as considerações explanatórias **conduza**, normalmente, a um conjunto manipulável de partições do espaço lógico ou, mesmo que isso fosse possível, isso não se deverá apenas ao fato de estarmos negligenciando um grande número de modos plausíveis, prima facie, de partição do espaço lógico (a última questão, claramente, ecoa no argumento do conjunto defeituoso de hipóteses).

Outra sugestão sobre a relação entre abdução e o raciocínio bayesiano, encontrado em Okasha (2000), McGrew (2003), e Lipton (2004, cap. 7), assinala que as considerações explanatórias podem servir como uma heurística para determinar, mesmo que apenas aproximadamente, probabilidades prévias e verossimilhança em casos nos quais, de outra forma, não teríamos nenhuma pista e não poderíamos ir além de meros palpites. Esta sugestão é sensível ao fato bem reconhecido de que nem sempre podemos atribuir uma probabilidade prévia a todas as hipóteses de nosso interesse, ou indicar o quão provável uma parte da evidência está condicionada a uma dada hipótese. Considerações acerca do poder explanatório dessa hipótese poderia, então, auxiliar na descoberta, talvez apenas dentro de certos limites, de quais probabilidades prévias devemos a ela atribuir, ou qual verossimilhança devemos a ela atribuir na posse das evidências disponíveis.

Os bayesianos, especialmente os mais modestos, talvez quisessem assinalar que o procedimento bayesiano deve ser seguido se, e somente se, ou (a) probabilidades prévias e verossimilhanças possam ser determinadas com algum grau de precisão e objetividade, ou (b) verossimilhanças possam ser determinadas com alguma precisão e, pode-se esperar que as probabilidades prévias possam ser "eliminadas" na medida em que se acumule mais e mais evidências, ou (c) pode-se esperar que ambas as probabilidades prévias e as verossimilhanças possam ser eliminadas. Nos casos restantes – eles poderiam dizer – devemos simplesmente evitar o emprego do raciocínio bayesiano. *A fortiori*, então, não há nesses casos a necessidade de um bayesianismo aprimorado pela abdução. Além disso, alguns resultados matemáticos incontestáveis indicam que, nos casos enquadrados em (a), (b), ou (c), as probabilidades, de qualquer forma, irão convergir para a verdade. Consequentemente, também não há nesses casos a necessidade de nenhum tipo de abdução heurística que os autores acima mencionados sugerem. A esse respeito,

Weisberg (2009, Seção 3.2) indica preocupações semelhantes.

Psillos (2000) ainda propõe outra forma em que a abdução poderia suplementar a teoria bayesiana da confirmação, uma que apresenta muito o espírito do conceito de abdução de Peirce. A ideia é que a abdução pode nos ajudar a selecionar possíveis candidatas para testagem, em que a testagem em questão é realizada em termos bayesianos. Psillos (2004), entretanto, entende que essa proposta atribui um papel à abdução que seria compreendida como sendo muito limitada pelos explicacionistas comprometidos.

Finalmente, uma possibilidade que não foi até agora considerada na literatura é a de que abdução e bayesianismo, juntos, não trazem muita relevância – como fazem nas propostas acima – quando considerados como diferentes formas de raciocínio; os bayesianos e os explicacionistas são personagens que aparecem em diferentes peças, por assim dizer. É amplamente aceito que, às vezes, pensamos e falamos sobre nossas crenças de modo categorial, enquanto outras vezes, pensamos e falamos sobre elas de modo mais gradual. Está longe de ser evidente como esses distintos modos de pensar e falar sobre crenças – a epistemologia da crença e a epistemologia dos graus de crença, para usar a terminologia de Richard Foley (1992) – estão relacionados. De fato, é uma questão em aberto se há alguma conexão direta entre as duas, ou mesmo se haveria uma tal conexão. Seja como for, dado que essa distinção é inegável, é uma sugestão plausível que, assim como há diferentes modos de falar e de pensar sobre crenças, há diferentes modos de falar e de pensar sobre a **revisão** de crenças. Em particular, a abdução pode muito bem ter seu lugar na epistemologia da crença, e ser evocada sempre que raciocinarmos sobre nossas crenças de forma categorial, enquanto, ao mesmo tempo, a regra de Bayes pode ter lugar na epistemologia dos graus de crença. Bayesianos intransigentes podem insistir que qualquer raciocínio que proceda de modo categorial pode eventualmente ser justificável em termos bayesianos, mas isso pressupõe a existência de princípios construídos que conectem a epistemologia da crença com a epistemologia dos graus de crença – e, como mencionado, atualmente não é claro se esses princípios existem.

Suplemento: Abdução peirciana

O termo “abdução” foi cunhado por Charles Sanders Peirce em sua obra sobre a lógica da ciência. Ele introduziu esse termo para denotar um tipo de inferência

não dedutiva que era diferente do já conhecido tipo de indução. É uma crítica comum que nenhum cenário coerente sobre o conceito de abdução emerja das obras de Peirce (embora isso, talvez, não seja surpreendente, dado que o autor desenvolveu o conceito de abdução ao longo de sua carreira, que durou mais de cinquenta anos). Para um relato conciso e ainda assim completo do desenvolvimento dos pensamentos de Peirce sobre abdução, consulte Fann (1970). No entanto, é claro que, segundo o entendimento de Peirce, o termo "abdução" não significa exatamente o que se entende por abdução atualmente (CAMPOS, 2011; MCAULIFFE, 2015). Uma diferença crucial entre a concepção de Peirce e o conceito moderno de abdução é que, de acordo com o último, abdução pertence ao que os lógicos empiristas denominam "contexto de justificação" – a fase da investigação científica em que se considera a avaliação de teorias – enquanto que para Peirce, a abdução tem seu devido lugar no contexto de descoberta, a fase da investigação científica em que se procura gerar novas teorias que podem, então, ser posteriormente avaliadas. Como ressalta Peirce "[a]bdução é o processo de formação de hipóteses explicativas. É a única operação lógica que introduz qualquer ideia nova (CP 5.172); em outro lugar, Peirce assinala que a abdução incorpora "[...] todas as operações das quais teorias e concepções são geradas" (CP 5.590). Dedução e indução, assim, desempenham um papel na fase posterior de avaliação teórica: a dedução auxilia na derivação de consequências testáveis das hipóteses explanatórias geradas a partir da abdução; a indução, por fim, auxilia na determinação de um veredito para as hipóteses, em que a natureza desse veredito depende do número de consequências testadas que foram verificadas (como parênteses, deve-se notar que Gerhard Schurz defendeu recentemente uma versão da abdução que, mais uma vez, incorpora o espírito peirciano. Segundo essa versão de abdução "[...] a função crucial de um padrão de abdução...consiste em sua função como uma estratégia de busca que nos leva, em um dado tipo de contexto, em um tempo razoável, à conjecturas explanatórias mais promissoras que são, assim, sujeitas ao teste futuro" (SCHURZ, 2008, p. 205). Esse artigo também é interessante pois desenvolve uma tipologia útil dos padrões de abdução).

Harry Frankfurt (1958) observou, entretanto, que a versão da abdução acima mencionada não é tão simples de compreender como pode parecer à primeira vista. Se pressupõe que a abdução constitua uma fase da lógica da ciência, mas o que é exatamente **lógico** sobre a geração de hipóteses explanatórias? De acordo com Peirce (CP 5.189), a abdução pertence à lógica pois ela pode ser caracterizada

pela seguinte estrutura:

> Um fato surpreendente, **C**, é observado;
> Mas se **A** fosse verdadeira, **C** deixaria de ser surpreendente;
> Então, há razões para suspeitar que **A** é verdadeira

Mas Frankfurt observa, com razão, que essa não é uma inferência que conduz a uma nova ideia. Afinal de contas, a nova ideia, a hipótese explanatória **A** – deve ocorrer a alguém **antes** que alguém infira que há razões para suspeitar que *A* seja verdadeira, pois *A* já está contida na segunda premissa.

Frankfurt argumenta, então, que em várias passagens na obra de Peirce, o autor sugere um entendimento da abdução não como um processo de *geração* de hipóteses, mas como um processo de **adoção** de hipóteses, em que a adoção de uma hipótese não diz respeito a hipótese ser verdadeira, verificada ou confirmada, mas ser uma candidata que merece ser investigada. Segundo esse entendimento, abdução ainda pode ser pensada como parte do contexto da descoberta. Ela desempenharia um tipo de função de seleção, ou filtro, determinando quais hipóteses concebidas na fase da descoberta estão aptas para a próxima fase, sendo sujeitas ao teste empírico. Como critério de seleção de hipóteses deve haver uma razão para suspeitar que a hipótese escolhida é verdadeira, e que teríamos tal razão se a hipótese explicasse qualquer fato observado que seja interessante, fazendo com que esses fatos deixem de ser surpreendentes. Essa proposta daria mais sentido à afirmação de Peirce segundo a qual a abdução é uma operação lógica.

Em última análise, entretanto, Frankfurt também rejeita essa proposta. Considerando, o autor ressalta, a possibilidade de haver infinitas hipóteses que expliquem um dado fato ou um conjunto de fatos, o que Peirce reconheceu, dificilmente seria uma condição suficiente para a adoção de uma hipótese (no sentido acima) que sua verdade faça com que um fato, ou um conjunto de fatos, deixe de ser surpreendente.

No mínimo, a abdução não parece ser muito útil como um critério para seleção. Entretanto, alguém poderia questionar se essa objeção seria válida. Ecoando no que dissemos sobre os argumentos de subdeterminação, notamos que de modo algum é claro se “explicar um dado fato” seja identificado com “fazer o fato deixar de ser surpreendente”. Segundo o entendimento de Frankfurt, para que uma hipótese explique um determinado fato, basta que ela implique tal fato. Mas,

virtualmente, nenhum filósofo da ciência nos dias de hoje entende que implicação é suficiente para explicação. E pareceria razoável ler a sentença "fazer com que um fato deixe de ser surpreendente" como "conceber uma explicação satisfatória para aquele fato". Em resposta à objeção de Frankfurt, pode-se argumentar que, mesmo se houver uma infinidade de hipóteses que expliquem um determinado fato, ainda sim haveria apenas uma pequena porção de hipóteses que se poderia afirmar que concebem uma explicação satisfatória ao fato em questão. Mas cabe aos estudiosos de Peirce, à luz de seus outros escritos, decidir se essa interpretação proposta é plausível.

Mesmo se "fazer com que um fato deixe de ser surpreendente" possa ser lido como "conceber uma explicação satisfatória para tal fato", é notável que não haja nenhuma referência nos escritos de Peirce à noção de melhor explicação. Algumas explicações satisfatórias ainda podem ser melhores que outras, e ainda poderia haver uma única **melhor** explicação. Essa ideia é crucial em todo pensamento recente sobre abdução. Aqui reside outra diferença principal entre a concepção peirciana de abdução e a concepção moderna.

Refeêrncias

ACHINSTEIN, P. **The Book of Evidence**. Oxford: Oxford University Press, 2001.

ADLER, J. Testimony, Trust, Knowing. **Journal of Philosophy**, v. 91, p. 264-275, 1994.

BACH, K.; HARNISH, R. **Linguistic Communication and Speech Acts.** Cambridge: MIT Press, 1979.

BIRD, A. **Philosophy of Science**. London: UCL Press, 1998.

BIGELOW, J. Quine, Mereology, and Inference to the Best Explanation. **Logique et Analyse**, v. 212, p. 465-482, 2010.

BOVENS, L.; HARTMANN, S. Solving the Riddle of Coherence. **Mind**, v. 112, p. 601-633, 2003.

BOYD, R. Scientific Realism and Naturalistic Epistemology. *In:* ASQUITH, P.; GIERE, R. (ed.). **East Lansing MI**: Philosophy of Science Association, 1980, p. 613-662. vol. II.

BOYD, R. The Current Status of Scientific Realism. *In:* LEPLIN, J. (ed.). **Scientific Realism.** Berkeley: University of California Press, 1984, p. 41-82.

BOYD, R. Lex Orandi est Lex Credendi. *In:* CHURCHLAND, P.; HOOKER, C. (ed.). **Images of Science.** Chicago: University of Chicago Press, 1985, p. 3-34.

BREM, S.; RIPS, L. J. Explanation and Evidence in Informal Argument. **Cognitive Science**, v. 24, p. 573-604, 2000.

CALLEBAUT, W. (ed.). **Taking the Naturalistic Turn.** Chicago: University of Chicago Press, 1993.

CAMPOS, D. On the Distinction Between Peirce's Abduction and Lipton's Inference to the Best Explanation. **Synthese**, v. 180, p. 419-442, 2011.

CLIMENHAGA, N., forthcoming. Inference to the Best Explanation Made Incoherent. **Journal of Philosophy**, v. 114, n. 5, 2017.

DASCAL, M. Conversational Relevance. *In:* MARGALIT, A. (ed.). **Meaning and Use.** Dordrecht: Reidel, 1979, p. 153-174.

DELLSÉN, F. Reactionary Responses to the Bad Lot Objection. **Studies in History and Philosophy of Science**, v. 61, p. 32-40, 2017.

DOUVEN, I. Inference to the Best Explanation Made Coherent. **Philosophy, of Science**, v. 66, p. S424-S435, 1999.

DOUVEN, I. Testing Inference to the Best Explanation. **Synthese**, v. 130, p. 355-377, 2002.

DOUVEN, I. Underdetermination. *In:* PSILLOS, S.; CURD, M. (ed.). **The Routledge Companion to the Philosophy of Science**. London: Routledge, 2008, p. 292-301.

DOUVEN, I. Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation. **Philosophical Quarterly**, v. 63, p. 428–444, 2013.

DOUVEN, I. **The Epistemology of Indicative Conditionals.** Cambridge: Cambridge University Press, 2016a.

DOUVEN, I. Explanation, Updating, and Accuracy. **Journal of Cognitive Psychology**, v. 28, p. 1004-1012, 2016b.

DOUVEN, I. What Is Inference to the Best Explanation? And Why Should We Care? *In:* POSTON, T.; Mccain, K. (ed.). **Best Explanations**: New Essays on Inference to the Best Explanation. Oxford: Oxford University Press, 2017.

DOUVEN, I.; SCHUPBACH, J. The Role of Explanatory Considerations in Updating. **Cognition**, v. 142, p. 299-311, 2015a.

DOUVEN, I.; SCHUPBACH, J. Probabilistic Alternatives to Bayesianism: The Case of Explanationism. **Frontiers in Psychology**, v. 6, p. 459, 2015b. DOI:10.3389/fpsyg.2015.00459.

DOUVEN, I.; WENMACKERS, S. Inference to the Best Explanation versus Bayes's Rule in a Social Setting. **British Journal for the Philosophy of Science**, 2015. DOI:10.1093/bjps/axv025.

DRAGULINSCU, S. Inference to the Best Explanation and Mechanisms in Medicine. **Theoretical Medicine and Bioethics**, v. 37, n. 3, p. 211-232, 2016.

DRAGULINSCU, S. Inference to the Best Explanation as a Theory for the Quality of Mechanistic Evidence in Medicine. **European Journal for Philosophy of Science,** 2016. DOI:10.1007/s13194-016-0165-x.

FANN, K. T. **Peirce's Theory of Abduction**. The Hague: Martinus Nijhoff, 1970.

FINE, A. The Natural Ontological Attitude. *In:* LEPLIN, J. (ed.). **Scientific Realism.** Berkeley: University of California Press, 1984, p. 83-107.

FOLEY, R. The Epistemology of Belief and the Epistemology of Degrees of Belief. **American Philosophical Quarterly**, v. 29, p.111-124, 1992.

FOLINA, J. Realism, Skepticism, and the Brain in a Vat. *In:* GOLDBERG, S. (ed.). **The Brain in a Vat.** Cambridge: Cambridge University Press, 2016, p. 155-173.

FORSTER, M.; SOBER, E. How to Tell when Simpler, More Unified, or Less Ad Hoc, Theories will Provide More Accurate Predictions. **British Journal for the Philosophy of Science**, v. 45, p. 1-36, 1994.

FRANKFURT, H. Peirce's Notion of Abduction. **Journal of Philosophy**, v. 55, p. 593-596, 1958.

FRICKER, E. Against Gullibility. *In:* MATILAL, B. K.; CHAKRABARTI, A. (ed.). **Knowing from Words.** Dordrecht: Kluwer, 1994, p. 125-161.

GOLDMAN, A. **Empirical Knowledge.** Berkeley: University of California Press, 1988.

HÁJEK, A. What Conditional Probability Could Not Be. **Synthese**, v.137, p. 273-323, 2003.

HARMAN, G. The Inference to the Best Explanation. **Philosophical Review**, v. 74, p. 88-95, 1965.

HARMAN, G. **Thought**. Princeton: Princeton University Press, 1973.

HARMAN, G. **Pragmatism and Reasons for Belief.** *In:* KULP, C. (ed.). **Realism/ Antirealism and Epistemology.** Totowa: Rowman and Littlefield, 1997, p. 123-147.

HARRÉ, R. **Varieties of Realism.** Oxford: Blackwell, 1986.

HARRÉ, R. Realism and Ontology. **Philosophia Naturalis**, v. 25, p. 386-398, 1988.

HOBBS, J. R. Abduction in Natural Language Understanding. *In:* HORN, L.; WARD, G. (ed.). **The Handbook of Pragmatics.** Oxford: Blackwell, 2004, p. 724-741.

JANSSEN, M. Reconsidering a Scientific Revolution: The Case of Einstein versus, Lorentz. **Physics in Perspective**, v. 4, p. 421- 446, 2002.

JOSEPHSON, J. R.; JOSEPHSON, S. G. (ed.). **Abductive Inference.** Cambridge: Cambridge University Press, 1994,

KITCHER, P. Real Realism: The Galilean Strategy. **Philosophical Review**, v. 110, p. 151-197, 2001.

KOEHLER, D. J. Explanation, Imagination, and Confidence in Judgment. **Psychological Bulletin**, v. 110, p. 499-519, 1991.

KOSLOWSKI, B.; MARASIA, J.; CHELENZA, M.; DUBLIN, R. Information Becomes Evidence when an Explanation Can Incorporate it into a Causal Framework. **Cognitive Development**, v. 23, p. 472-487, 2008.

KOSSO, P. **Reading the Book of Nature**. Cambridge: Cambridge University Press, 1992.

KRZYŻANOWSKA, K.; WENMACKERS, S.; DOUVEN, I. Rethinking Gibbard's Riverboat Argument. **Studia Logica**, v. 102, p. 771-792, 2014.

KUIPERS, T. Approaching the Truth with the Rule of Success. **Philosophia Naturalis**, v. 21, p. 244-253, 1984.

KUIPERS, T. Naive and Refined Truth Approximation. **Synthese**, v. 93, p. 299-341, 1992.

KUIPERS, T. **From Instrumentalism to Constructive Realism**. Dordrecht: Kluwer, 2000.

KVANVIG, J. A Critique of van Fraassen's Voluntaristic Epistemology. **Synthese**, v. 98, p. 325-348, 1994.

KYBURG Jr. H. **Science and Reason.** Oxford: Oxford University Press, 1990.

LAUDAN, L. A Confutation of Convergent Realism. **Philosophy of Science**, v. 48, p. 19-49, 1981.

LEWIS, D. A Subjectivist's Guide to Objective Chance. *In:* JEFFREY, R. (ed.). **Studies in Inductive Logic and Probability.** Berkeley: University of California Press, 1980. p. 263-293.

LI, M.; VITANYI, P. An Introduction to Kolmogorov Complexity and its Applications. New York: Springer, 1997.

LIPTON, P. **Inference to the Best Explanation**. London: Routledge, 1991.

LIPTON, P. Is the Best Good Enough? **Proceedings of the Aristotelian Society**, v. 93, p. 89-104, 1993.

LIPTON, P. The Epistemology of Testimony. **Studies in History and Philosophy of Science**, v. 29, p. 1-31, 1998.

LIPTON, P. **Inference to the Best Explanation**. 2. ed. London: Routledge, 2004.

LOMBROZO, T. Simplicity and Probability in Causal Explanation. **Cognitive Psychology**, v. 55, p. 232-257, 2007.

LOMBROZO, T. Explanation and Abductive Inference. *In:* HOLYOAK, K.; MORRISON, R. (ed.). **Oxford Handbook of Thinking and Reasoning.** Oxford: Oxford University Press, 2012, p. 260-276.

LOMBROZO, T. Explanatory Preferences Shape Learning and Inference. **Trends in Cognitive Sciences**,v. 20, p. 748-759, 2016.

LOMBROZO, T.; GWYNNE, N. Z. Explanation and Inference: Mechanistic and Functional Explanations Guide Property Generalization. **Frontiers in Human Neuroscience**, v. 8, 2014. DOI:10.3389/fnhum.2014.00700.

MAHER, P. Diachronic Rationality. **Philosophy of Science**, v. 59, p. 120-141, 1992.

MCAULIFFE, W. How Did Abduction Get Confused with Inference to the Best Explanation? Transactions of the Charles S. Peirce. **Society**, v. 51, p. 300-319, 2015.

MCCAIN, K.; POSTON, T. Why Explanatoriness is Evidentially Relevant. **Thought**, v. 3, p. 145-153, 2014.

MCGREW, T. Confirmation, Heuristics, and Explanatory Reasoning. **British Journal for the Philosophy of Science**, v. 54, p. 553-567, 2003.

MCMULLIN, E. **The Inference that Makes Science.** Milwaukee: Marquette University Press, 1992.

MCMULLIN, E. Epistemic Virtue and Theory Appraisal. *In:* DOUVEN, I.; HORSTEN, L. (ed.). **Realism in the Sciences.** Leuven: Leuven University Press, 1994, p. 13-34.

MOORE, G. E. **Proof of an External World in his Philosophical Papers**. New York: Collier Books, 1962, p. 126-149.

MOSER, P. **Knowledge and Evidence.** Cambridge: Cambridge University Press, 1989.

MUSGRAVE, A. The Ultimate Argument for Scientific Realism. *In:* NOLA, R. (ed.). **Relativism and Realism in Science.** Dordrecht: Kluwer, 1988, p. 229-252.

NIINILUOTO, I. Verisimilitude: The Third Period. **British Journal for the Philosophy of Science**, v. 49, p. 1-29, 1998.

NIINILUOTO, I. Defending Abduction. **Philosophy of Science**, v. 66,p. S436-S451, 1999.

OKASHA, S. Van Fraassen's Critique of Inference to the Best Explanation. **Studies in History and Philosophy of Science**, v. 31, p. 691-710, 2000.

OLSSON, E. **Against Coherence.** Oxford: Oxford University Press, 2005.

PARGETTER, R. The Scientific Inference to Other Minds. **Australasian Journal of Philosophy**, v. 62, p. 158-163, 1984.

PEIRCE, C. S. **Collected Papers of Charles Sanders Peirce**. Cambridge: Harvard University Press, 1931–1958.

POSTON, T. **Reason and Explanation.** Basingstoke: Palgrave Macmillan, 2014.

PSILLOS, S. **Scientific Realism**: How Science Tracks Truth. London: Routledge, 1999.

PSILLOS, S. Abduction: Between Conceptual Richness and Computational Complexity. *In:* KAKAS, A. K.; FLACH, P. (ed.). **Abduction and Induction**: Essays on their Relation and Integration. Dordrecht: Kluwer, 2000, p. 59-74.

PSILLOS, S. Inference to the Best Explanation and Bayesianism. *In:* F. Stadler (ed.). **Induction and Deduction in the Sciences.** Dordrecht: Kluwer, 2004, p. 83-91.

PUTNAM, H. **Reason, Truth and History**. Cambridge: Cambridge University Press, 1981.

ROCHE, W.; SOBER, E. Explanatoriness is Evidentially Irrelevant, or Inference to the Best Explanation Meets Bayesian Confirmation Theory. **Analysis**, v. 73, p. 659-668, 2013.

ROCHE, W.; SOBER, E. Explanatoriness and Evidence: A Reply to McCain and Poston. **Thought**, v. 3, p. 193-199, 2014.

RUSSELL, B. **The Problems of Philosophy.** Oxford: Oxford University Press, 1912.

SCHUPBACH, J. Is the Bad Lot Objection Just Misguided? **Erkenntnis**, v. 79, p. 55-64, 2014.

SCHUPBACH, J.; SPRENGER, J. The Logic of Explanatory Power. **Philosophy of Science**, v. 78, p. 105-127, 2011.

SCHURZ, G. Patterns of Abduction. **Synthese**, v. 164, p. 201-234, 2008.

SHALKOWSKI, S. IBE, GMR, and Metaphysical Projects. *In:* HALE, B.; HOFFMANN, A. (ed.). **Modality: Metaphysics, Logic, and Epistemology.** Oxford: Oxford University Press, 2010, p. 169-187.

SKYRMS, B. A Mistake in Dynamic Coherence Arguments?" **Philosophy of Science**, v. 60, p. 320-328, 1993.

SLOMAN, S. When Explanations Compete: The Role of Explanatory Coherence on Judgments of Likelihood. **Cognition**, v. 52, p. 1-21, 1994.

SOBER, E. **Ockham's Razor**: A User's Manual, Cambridge: Cambridge University Press, 2015.

STANFORD, K. Underdetermination of Scientific Theory. *In:* ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**, 2009. Disponível em: <https://plato.stanford.edu/archives/win2009/entries/scientific-underdetermination>.

Acesso em: 27 set. 2021.

TELLER, P. Conditionalization and Observation. **Synthese**, v. 26, p. 218-258, 1973.

THAGARD, P. The Best Explanation: Criteria for Theory Choice. **Journal of Philosophy**, v. 75, p. 76-92, 1978.

VAN FRAASSEN, B. **The Scientific Image.** Oxford: Oxford University Press, 1980.

VAN FRAASSEN, B. Glymour on Evidence and Explanation. *In:* EARMAN, J. (ed.). **Testing Scientific Theories.** Minneapolis: University of Minnesota Press, 1983, p. 165-176.

VAN FRAASSEN, B. Empiricism in the Philosophy of Science. *In:* CHURCHLAND, P.; HOOKER, C. (ed.). **Images of Science.** Chicago: University of Chicago Press, 1985, p. 245-308.

VAN FRAASSEN, B. **Laws and Symmetry.** Oxford: Oxford University Press, 1989.

VOGEL, J. Cartesian Skepticism and Inference to the Best Explanation. **Journal of Philosophy**, v. 87, p. 658-666, 1990.

VOGEL, J. The Refutation of Skepticism. *In:* STEUP, M.; SOSA, E. (ed.). **Contemporary Debates in Epistemology.** Oxford: Blackwell Publishing, 2005, p. 72-84.

WEINTRAUB, R. Induction and Inference to the Best Explanation. **Philosophical Studies**, v. 166, p. 203-216, 2013.

WEISBERG, J. Locating IBE in the Bayesian Framework. **Synthese**, v. 167, p. 125-143, 2009.

WILLIAMS, J.; LOMBROZO, T. The Role of Explanation in Discovery and Generalization: Evidence from Category Learning. **Cognitive Science**, v. 34, p. 776-806, 2010.

WILLIAMSON, T. Semantic Paradoxes and Abductive Methodology. *In:* ARMOUR-GARB, B. (ed.). **The Revenge of the Liar.** Oxford: Oxford University Press, 2017. DOI: 10.1093/oso/9780199896042.003.0013.

(IV) PREDIÇÃO VERSUS ACOMODAÇÃO*

Autor: Eric Barnes
Tradução: Tiago L. T. de Oliveira
Revisão: Ana Margarete Barbosa de Freitas

Na literatura filosófica inicial, uma "predição" era considerada como uma consequência empírica de uma teoria que ainda não havia sido verificada no tempo em que estava sendo construída - uma 'acomodação' era aquela que havia passado pela verificação. A visão de que predições são superiores às acomodações na avaliação de teorias científicas é conhecida como 'preditivismo'. Comumente, entretanto, o preditivismo é entendido, mais precisamente, como a implicação de que a evidência confirma mais fortemente a teoria quando prevista do que acomodada. Muita tinta foi derramada modificando o conceito de 'predição' e explicando por que o preditivismo é ou não é verdadeiro, e se a história da ciência revela que os cientistas são preditivistas em suas avaliações de teorias. O debate sobre o

---

* BARNES, E. Prediction versus Accommodation. *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/prediction-accommodation/. Acesso em: 18 set. 2021.

The following is the translation of the entry on Prediction versus Accommodation by Eric Barnes, 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/prediction-accommodation/. 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/ prediction-accommodation /. 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.

preditivismo também se revela de modo importante no debate sobre o realismo científico.

1 Introdução Histórica

Nos séculos XVIII e XIX houve um acalorado debate sobre o método científico - num lado do ringue estava o 'método da hipótese' que postulava hipóteses sobre entidades inobserváveis que 'salvavam os fenômenos' e, assim, seriam indiscutivelmente verdadeiras (ver Laudan [1981a]). Críticos desse método apontavam que hipóteses poderiam sempre ser ajustadas artificialmente para acomodar qualquer quantidade de dados. Mas foi notado que algumas dessas teorias tinham a virtude adicional de gerar predições específicas de fenômenos até então não observados - assim, cientistas como John Herschel e William Whewell argumentaram que hipóteses que salvam fenômenos poderiam ser justificadas quando fossem confirmadas por tais 'novos' fenômenos. Whewell sustentou que predições carregam peso especial porque uma teoria que corretamente prediz um resultado surpreendente não pode fazê-lo por acaso e, portanto, deve ser verdadeira (Whewell,1849 [1968, p. 294]). Pareceu assim que a evidência predita confirmava a teoria de modo mais forte do que a evidência acomodada. Mas John Stuart Mill (em seu debate com Whewell) categoricamente negou aquela alegação, afirmando que

> tais predições e seus requisitos são, de fato, bem calculadas para impressionar o vulgar ignorante, cuja fé na ciência repousa unicamente sobre coincidências similares entre suas profecias e o que vem a acontecer. Mas é estranho que qualquer pressão considerável seja depositada em tal coincidência por pensadores científicos (1843, v. 2, p. 23).

John Maynard Keynes fornece uma explicação simples do porquê de o preditivismo possuir uma enganadora aparência de verdade em uma breve passagem de seu livro *A Treatise on Probability* [Um tratado sobre probabilidade]:

> A virtude peculiar da predição ou da pré-designação é de todo imaginária... A plausibilidade do argumento [para o preditivismo] é derivada de uma fonte diferente. Se uma hipótese é proposta a *priori*, isso comumente

> significa que há algum fundamento para ela, decorrente de nosso conhecimento prévio, separada de uma base puramente indutiva e, se tal é o caso, a hipótese é claramente mais forte que aquela que repousa em base indutiva apenas. Mas se ela for meramente um palpite, o fato de sorte de que a hipótese precedeu alguns ou todos os casos que a verificam não acrescenta nada que seja a seu valor. É a união entre conhecimento anterior e fundamento indutivo que surge de instâncias imediatas, que confere peso a qualquer hipótese, e não a ocasião na qual a hipótese é primeiro proposta (1921, p. 305-306)[34] .

Por 'fundamento indutivo' para uma hipótese, Keynes claramente quer dizer os dados que a hipótese encaixa. Keynes quer dizer que, quando algum teórico que se compromete a testar uma hipótese a propõe pela primeira vez, tipicamente alguma outra forma (presumivelmente teórica) de apoio estimulou a proposta. Dessa forma, hipóteses que são propostas sem serem construídas para encaixar os dados empíricos (que subsequentemente elas mostrarão implicar) são tipicamente mais bem apoiadas do que hipóteses propostas apenas para encaixar os dados - pois as últimas carecem do apoio independente possuído pelas primeiras. A aparência de plausibilidade do preditivismo emerge porque o papel das evidências indutoras de hipóteses preliminares está sendo suprimido.

Karl Popper é provavelmente o mais famoso proponente da predição na história da filosofia. Em sua palestra *Science: conjectures and refutations* [Ciência: conjecturas e refutações] Popper reconta sua tentativa de juventude de lidar com a questão: "Quando uma teoria deveria ser classificada como científica?" (POPPER, 1963, p. 33-65). Popper ficou convencido de que certas teorias populares do seu tempo, incluindo a teoria da história de Marx e a psicanálise freudiana, eram pseudociências. Popper considerou o problema de distinguir entre teorias científicas e pseudocientíficas como 'o problema da demarcação'. Sua solução para o problema

---

[34] Citando esta passagem, Achinstein (1994, p. 163; 2001, cap. 10) argumenta que a aparência de plausibilidade do preditivismo é ilusória, e pode ser explicada ao citar outros fatores, tais como o procedimento de seleção usado para gerar a evidência. Howson articula uma 'dissolução keynesiana' do preditivismo similar, confira a Seção 7.

da demarcação, como é bem sabido, foi identificar a qualidade de falseabilidade (ou 'testabilidade') como a marca da teoria científica.

As pseudociências eram marcadas, alegou Popper, por seu vasto poder explicativo. Elas podiam explicar não somente todos os fenômenos reais relevantes que o mundo apresentava, mas também quaisquer fenômenos concebíveis que pudessem ser enquadrados em seu domínio. Isso ocorria porque as explicações oferecidas pelas pseudociências eram suficientemente maleáveis, de modo a sempre poderem ser ajustadas *ex post facto* para explicar qualquer coisa. Assim, as pseudociências nunca correram o risco de ser inconsistentes com os dados. Em contraste, uma teoria genuinamente científica fazia previsões específicas sobre o que deveria ser observado e, portanto, corria o risco de falsificação. Popper enfatizou que aquilo que estabelecia o caráter científico da teoria da relatividade era que ela 'colocava seu pescoço à prova' de um modo que as pseudociências nunca fizeram.

Como Whewell e Herschel, Popper apela para as predições que uma teoria faz como um modo de separar os usos ilegítimos do método da hipótese de seus usos legítimos. Mas enquanto Whewell e Herschel apontavam o sucesso preditivo como condição necessária para a aceitabilidade de uma teoria que fora gerada pelo método da hipótese, Popper, em sua solução para o problema da demarcação, focou não no sucesso de uma predição, mas no fato de que, no todo, a teoria fez uma predição. Havia para Popper, é claro, uma importante diferença entre teorias científicas cujas predições foram confirmadas e aquelas cujas predições foram falsificadas. Teorias falsificadas eram para ser rejeitadas, ao passo que teorias que sobreviveram aos testes eram para ser 'provisoriamente aceitas' até serem falsificadas. Popper não sustentou, como Whewell e Herschel, que predições bem-sucedidas poderiam constituir prova legítima de uma teoria, na verdade, Popper sustentou que era impossível mostrar que uma teoria fosse sequer provável baseado na evidência, pois ele abraçou a crítica de Hume à lógica indutiva, que tornava impossível o apoio evidencial para a verdade de teorias. Assim, pode-se atribuir a Popper um comprometimento com o preditivismo somente no sentido amplo de que ele sustentou que predições seriam superiores a acomodações - ele não sustentou que predições confirmavam teorias de modo mais forte do que acomodações. Reconciliar a afirmação de Popper de que uma teoria que desfrutava de sucesso preditivo deveria ser 'provisoriamente aceita' com seu anti-indutivismo se provaria definitivamente impossível (SALMON, 1981).

Imre Lakatos (1970, 1971) propôs uma explicação do método científico na forma de sua 'metodologia dos programas de pesquisa científica' que era um

desenvolvimento da abordagem de Popper. Um programa de pesquisa científica era constituído por um 'núcleo duro' de proposições que eram retidas ao longo da vida daquele programa, em união com um 'cinturão protetor' que era constituído por hipóteses auxiliares que eram ajustadas de modo a reconciliar o núcleo duro com os dados empíricos. A tentativa de partes dos proponentes do programa de pesquisa de reconciliar o programa com os dados empíricos produzia uma série de teorias $T_1, T_2, \ldots T_n$ onde, pelo menos em alguns casos, T_{i+1} serve para explicar alguns dados anômalos para T_i. Lakatos sustentou que um programa de pesquisa era 'teoricamente progressivo' na medida em que cada nova teoria prediz algum fato novo até então inesperado. Um programa de pesquisa é 'empiricamente progressivo' na medida em que seu novo conteúdo empírico estava corroborado, ou seja, se cada nova teoria leva à descoberta de "algum novo fato" (LAKATOS, 1970, p. 118). Lakatos assim ofereceu uma nova solução para o problema da demarcação: um programa de pesquisa era pseudocientífico na medida em que não seria teoricamente progressivo. A avaliação da teoria é construída em termos de competição de programas de pesquisa: um programa de pesquisa derrota um programa rival ao se mostrar mais progressiva empiricamente a longo prazo.

2. Hipóteses *Ad Hoc*

De acordo com o Dicionário *Merriam-Webster´s Collegiate*, alguma coisa é *ad hoc* se for "formada ou usada para necessidades ou problemas específicos ou imediatos". Uma hipótese *ad hoc* então é aquela formada para resolver um problema específico, tal como o problema de imunizar uma teoria em particular da falsificação por dados anômalos (e dessa forma acomodar aqueles dados). Consequentemente, o que torna uma hipótese *ad hoc*, no sentido do português ordinário, não tem nada a ver com o conteúdo de uma hipótese, mas simplesmente com a motivação do cientista que a propôs, e não é claro por que haveria qualquer coisa suspeita sobre tal motivação. Não obstante, hipóteses *ad hoc* tem sido, há tempos, suspeitas nas discussões sobre método científico. Uma suspeita que ressoa com o ceticismo preditivista acerca da acomodação.

Para Popper, uma conjectura é *ad hoc* "se ela é introduzida [...] para explicar uma dificuldade particular, mas [...] não pode ser independentemente testada" (POPPER, 1974, p. 986). Assim, a concepção de *ad hoccidade* de Popper adicionou ao significado ordinário do português um requisito a mais, no caso de uma hipótese

ad hoc que foi simplesmente introduzida para explicar um fenômeno singular, a hipótese *ad hoc* não possui outras consequências testáveis além daquele fenômeno. No caso de uma modificação de teoria *ad hoc* introduzida para resolver uma anomalia para a teoria, a teoria modificada não possui consequências testáveis além daquelas da teoria original.

Popper ofereceu duas explicações do porquê de as hipóteses *ad hoc* serem suspeitas. Uma delas era a de que se oferecermos **T** como explicação de **f**, mas então citarmos **f** como a única razão que possuímos para acreditar em **T**, Popper declara que nos empenhamos em raciocínio suspeito por razões de circularidade (POPPER, 1972, p. 192-3). Isso é indiscutivelmente falacioso da parte de Popper, uma prova circular ofereceria uma proposição, **p**, em apoio a uma segunda proposição **q**, quando **q** já havia sido dada como apoio de **p**. Mas no exemplo acima, enquanto **f** é dado como evidência para **T**, **T** é dado como explicação de (e não evidência para) **f**, e, assim, não há raciocínio circular (BAMFORD, 1993, p. 338).

A outra explicação de o porquê hipóteses *ad hoc* fossem consideradas com suspeição era o fato de elas se chocarem com a meta da ciência, o que, para Popper, incluía a proposta de teorias com acréscimo de conteúdo empírico, nomeadamente, acréscimo de falseabilidade. Hipóteses *ad hoc*, para Popper, sofrem de uma falta de testabilidade independente e assim reduzem (ou pelo menos falham em aumentar) a testabilidade das teorias que elas modificam (confira acima). Entretanto, a afirmação de Popper de que o processo de modificação de uma teoria *ad hoc* tende a levar a falseabilidade insuficiente e é 'prática não científica' tem sido desafiada (BAMFORD, 1993, p. 350).

Autores subsequentes argumentaram que uma hipótese proposta pelo bem de imunizar uma teoria contra falsificação poderia ser 'suspeita' por várias razões, e assim poderia ser *ad hoc* de várias maneiras. Zahar (1973) argumentou que uma hipótese era *ad hoc*$_1$ se ela não tivesse novas consequências em comparação com sua predecessora (isto é, não fosse testável independentemente), *ad hoc*$_2$ se nenhuma de suas novas predições tivessem sido realmente verificadas (quer porque não foi ainda testada quer por ter sido falsificada), e *ad hoc*$_3$

> se ela é obtida de sua predecessora através de uma modificação das hipóteses auxiliares que não concordam com o espírito da heurística do programa (1973, p. 101).

Além, então, do critério de Popper de uma falta de testabilidade independente, uma hipótese introduzida para acomodar algum dado poderia ser *ad hoc* porque ela não seria simplesmente confirmada (*ad hoc*$_2$) ou porque ela falhou em aderir aos compromissos básicos do programa de pesquisa no quais ela foi proposta (*ad hoc*$_3$).

Outra abordagem propõe que uma hipótese **H** introduzida numa teoria **T** em resposta a um resultado experimental **E** é *ad hoc* se ela não for amplamente suportada e parece ser uma tentativa superficial de encobrir problemas profundos com uma teoria que necessita realmente de revisão substancial. Assim, apontar a acusação de *ad hoccidade* contra uma hipótese era, na verdade, dirigir sério ceticismo à teoria que a hipótese visava salvar. Esse conceito de *ad hoccidade* certamente dá sentido à crítica de Einstein de considerar *ad hoc* a hipótese da contração de Lorentz-Fitzgerald, como hipótese suplementar à teoria do éter; e a postulação de Pauli de que o neutrino era um salvamento *ad hoc* da mecânica quântica clássica (LEPLIN, 1975, 1982). Para uma discussão mais aprofundada, consulte Grünbaum (1976).

Parece claramente verdadeiro que o julgamento da comunidade científica sobre se uma hipótese é *ad hoc* pode mudar. Dada essa revisibilidade, e a dimensão estética da avaliação de teorias (que deixa a avaliação em certo grau 'no olho de quem vê') pode não haver um ponto particular em aderir a uma teoria da *ad hoccidade*, se pelo termo *ad hoc* queremos dizer 'proposto ilegitimamente' (Hunt 2012).

3. Caracterizações Iniciais de Novidade

Popper escreveu que

> Confirmações deveriam contar apenas se elas são o resultado de predições arriscadas; quer dizer, se, sem a iluminação da teoria em questão, deveríamos ter esperado um evento que fosse incompatível com a teoria - um evento que teria refutado a teoria (1963, p. 36)

Dessa forma, Popper (e, subsequentemente, Lakatos) endossou uma condição temporal de novidade, uma predição conta como nova se ela não é sabidamente verdadeira (ou é esperado que se prove falsa) no momento em que a teoria é construída. Mas era bastante óbvio que isso tornaria implausíveis importantes questões de confirmação no momento em que certos fatos fossem conhecidos.

Assim Zahar propôs que um fato é novo "se ele não pertencia à situação-problema que governou a construção da hipótese" (1973, p. 103). Essa forma de novidade foi considerada 'novidade-problema' (GARDNER, 1982, p. 2). Mas no mesmo artigo Zahar se propõe a exemplificar sua concepção de novidade ao remeter ao caso em que Einstein não usou do conhecido comportamento do periélio de Mercúrio na construção de sua teoria da relatividade[35] . Gardner nota que essa última concepção de novidade, à qual chamou de 'novidade-uso', é distinta da novidade-problema (GARDNER, 1982, p. 3). A evidência é novidade-uso para **T** se **T** não foi construída para encaixar a evidência (quer ela fosse ou não parte da 'situação-problema' relevante que a teoria visava resolver). Na literatura subsequente, a assim chamada concepção heurística de novidade foi identificada com novidade-uso, ela foi articulada mais adiante em Worral (1978, 1985).[36]

Outra abordagem defende que uma consequência nova de uma teoria é a que não era conhecida pela teórica no momento em que ela formulou a teoria, parece-se com uma versão da concepção temporal, mas esse ponto apela implicitamente à concepção heurística: se uma teórica sabia de um resultado antes de construir uma teoria que o explica, pode ser difícil determinar se aquela teórica de algum modo adequou a teoria para se ajustar ao fato (ela pode ter feito isso inconscientemente). Uma concepção baseada em conhecimento é, assim, o melhor que podemos fazer para lidar com essa dificuldade (GARDNER, 1982).[37]

A concepção heurística é, todavia, profundamente controversa - porque ela torna a avaliação epistêmica de teorias curiosamente dependente da vida mental de seus construtores, especificamente sobre o conhecimento e as intenções do

---

[35] Thomason (1992) é uma avaliação detalhada e cética da heurística lakatosiana, empregando a concepção de novidade de Zahar, aplicada ao programa de pesquisa copernicano.

[36] Worrall representa sua explicação subsequente do preditivismo, descrita na seção 6.5 abaixo, como um desesenvolvimento da noção heurística que ele esboçou nesses artigos iniciais.

[37] Ao usar conceitos a serem explicados na próxima seção, Gardner, assim, endossa um preditivismo forte a respeito da novidade heurística e um preditivismo fraco a respeito da novidade baseada em conhecimento (e por implicação novidade temporal), como novidade baseada conhecimento (e assim novidade temporal) é evidência conclusiva da novidade heurística.

teórico para construir uma teoria que acomodou certos dados mais do que outros. O comentário de Leplin é típico:

> As esperanças do teórico, expectativas, conhecimento, intenções, ou o que quer que seja, não parecem se relacionar com a sustentação epistêmica de sua teoria de um modo que possa garantir a eles um papel pivotal (1997, p. 54).

Para comentários similares, confira Gardner (1982, p. 6), Thomason (1992, p. 195), Schlesinger (1987, p. 33), Achinstein (2001, p. 210-230) e Collins (1994).

Outra abordagem nota que cientistas trabalham com teorias em competição e que o papel das novas confirmações é o de decidir entre tais teorias. Assim, uma consequência de uma teoria **T** é uma 'nova predição' se ela não for uma consequência da melhor teoria disponível presente realmente na área para além de **T** (a predição do periélio de Mercúrio pela teoria da relatividade de Einstein constituiu uma nova predição porque ela não era, diretamente, uma consequência da mecânica newtoniana) (MUSGRAVE, 1974, p. 18). Trabalhando em um quadro lakatosiano, Frankel alega que uma consequência é nova, em relação a uma teoria e seu programa de pesquisa, se ela não for similar a um fato que já fora usado por membros do mesmo programa de pesquisa para apoiar uma teoria projetada para resolver os mesmos problemas que a teoria em questão (1979, p. 25). Também num quadro lakatosiano, Nunan alega que uma consequência é nova se não foi ainda usada para apoiar uma teoria apreciada em algum programa de pesquisa rival, ou se não pode prontamente ser explicada em termos de uma tal teoria. (1984, p. 279).[38]

Há claramente múltiplas formas de novidade e é, em geral, reconhecido que um fato poderia ser 'novo' em múltiplos sentidos, como veremos, alguns carregam mais peso epistêmico do que outros (MURPHY, 1989).

---

[38] A concepção de Leplin (1997) de novidade (confira Seção 8 abaixo) inclui a condição de Singularidade, que é uma instância de tal noção comparativa de novidade. Harker (2010) implanta tal concepção comparativa de novidade em sua articulação de um argumento para o realismo científico seletivo.

4. Uma Taxonomia Preditivista

O preditivismo global sustenta que predições são sempre superiores a acomodações, enquanto o preditivismo local sustenta que isso perdura apenas em certos casos. O preditivismo forte declara que a predição é instrinsicamente superior à acomodação, enquanto o preditivismo fraco sustenta que o sucesso preditivo é epistemicamente relevante porque ele é sintomático de outras características que possuem importância epistêmica. A distinção entre preditivismo forte e fraco permite combinações com distinções entre diferentes tipos de novidade. Por exemplo, poderia-se manter que predições temporais são intrinsecamente melhores que acomodações temporais (preditivismo temporal forte) ou que predições temporais fossem sintomáticas de alguma outra característica de boa construção de teorias. Acerca do preditivismo temporal fraco, confira Hitchcock e Sober (2004, p. 3-4). Essas distinções serão mais ilustradas a seguir.

5. A Tese do Suporte Nulo

Uma versão do preditivismo heurístico global forte é a tese do suporte nulo segundo a qual teorias nunca recebem confirmação da evidência que foram construídas para acomodar - precisamente pelo modo como foram construídas. Essa tese foi atribuída a Bacon e Descartes (HOWSON, 1990, p. 225). Popper e Lakatos também se filiam a essa tese, embora seja importante lembrar que eles não reconhecem qualquer forma de suporte confirmacional - mesmo de predições bem-sucedidas. Mas outros que mantiveram que predições bem-sucedidas confirmam teorias endossaram, apesar disso, a hipótese do suporte nulo. Giere oferece o seguinte argumento:

> Se os fatos conhecidos foram usados para construir o modelo e foram, assim, inseridos na hipótese resultante...então o ajuste entre esses fatos e a hipótese não fornece evidência para a hipótese ser verdadeira, [uma vez que] esses fatos não possuem chance de refutar a hipótese (1984, p. 161). Os autores Glymour (1980, p. 114) e Zahar (1983, p. 245) oferecem argumentos similares.

A ideia é que o modo como a teoria fora construída forneceu uma proteção ilegítima contra a falsificação pelos fatos, consequentemente os fatos não podem apoiar a teoria. Outros, entretanto, pensam que este argumento é ilusório, notando que, uma vez que o conteúdo da hipótese é fixado, não faz sentido pensar que quaisquer fatos tenham uma 'chance' de falsificar a teoria. A teoria diz o que diz, e qualquer fato particular a refuta ou não.

> Giere confundiu o que é de fato uma variável aleatória (a configuração experimental ou fonte de dados **E** juntamente com seu conjunto de distintos resultados possíveis) com um de seus valores (o resultado **e**) [...] Além disso, faz perfeitamente todo sentido dizer que **E** pôde bem ter produzido um resultado diferente do que *e*, que foi o que de fato foi produzido (HOWSON 1990, p. 229). Consulte, também, Collins (1994, p. 220).

Assim, o argumento de Giere colapsa.

Howson defendeu em uma série de artigos (1984, 1988, 1990) que a tese do suporte nulo é falsificada com o uso de exemplos simples, tais como os seguintes:

> Uma urna contém um número desconhecido de bilhetes pretos e brancos, onde a proporção **p** de bilhetes pretos é também desconhecida. Os dados consistem simplesmente em um relatório da frequência relativa **r** / **k** de bilhetes pretos em um número grande **k** de impressões com reposição pela urna. À luz dos dados propomos a hipótese de que **p** = *(***r** / **k**) + ϵ para algum ϵ adequado dependendo de **k**. Essa hipótese é, de acordo com a tradição estatística, muito bem apoiada pelos dados a partir dos quais ela é claramente construída (1990, p. 231).

Neste caso há, nota Howson, uma teoria de fundo que provê um modelo de experimento (isto é, uma sequência de ensaios de Bernoulli, ou seja, uma sequência de tentativas com dois resultados nos quais a probabilidade de obter cada resultado é a mesma em cada tentativa; ela deixa apenas um único parâmetro

para ser avaliado). Na medida em que possuímos boas razões para acreditar que esse modelo se aplica, nossa inferência para a alta probabilidade da hipótese é questão de metodologia estatística padrão, e a tese do suporte nulo está refutada.

Foi defendido que uma das limitações do bayesianismo é que ele é fatalmente comprometido com a, claramente falsa, tese do suporte nulo (GLYMOUR, 1980). A condição bayesiana padrão pela qual a evidência e apoia h é dada pela inequação **p(h | e) > p(h)**. Mas aqui é conhecida (e assim **p(e)=1**), temos **p(h | e) = p(h)**. Isso veio a ser conhecido como o "problema bayesiano da evidência antiga". Howson (1984) notou que esse problema poderia ser superado ao selecionar uma função de probabilidade **p*** baseada na afirmação de que e não era conhecida - assim, mesmo se **p(h | e) = p(h)**, ainda poderia sustentar que **p*(h | e) > p*(h)**. Deste modo seguiu-se uma extensa literatura sobre o problema da evidência antiga que não será resumida aqui (*vide* CHRISTIANSEN, 1999; EELLS; FITELSON, 2000; BARNES, 1999, p. 208, cap. 7; HARTNANN; FITELSON, 2015).

6. Teorias Contemporâneas do Preditivismo

6.1 Métodos de Descoberta Confiáveis

Patrick Maher (1988, 1990, 1993) apresentou um experimento de pensamento seminal e uma análise bayesiana de suas implicações preditivistas.

O experimento de pensamento continha dois cenários: no primeiro cenário, um sujeito (o acomodador) é apresentado a **E**, uma sequência de 99 arremessos de moeda. **E** forma uma sequência aparentemente aleatória de caras e coroas. O acomodador é então instruído a dizer-nos o resultado dos primeiros 100 arremessos, ele responde recitando **E** e, então, somando a predição de que a 100ª jogada será cara, a conjunção de **E** e essa última jogada é **T**. No outro cenário, outro sujeito (o preditor) é solicitado a predizer os primeiros 100 arremessos sem ter testemunhado qualquer resultado - o preditor endossa a teoria **T**. Depois disso a moeda é lançada 99 vezes, **E** é estabelecida e as primeiras 99 predições são confirmadas. A questão é saber em qual desses dois cenários **T** está melhor confirmada. É fortemente intuitivo que **T** está mais bem-confirmada no cenário do preditor do que no cenário do acomodador, sugerindo que o preditivismo detém a verdade neste caso. Se permitirmos que '*O*' seja a afirmação de que a evidência **E** foi inserida na construção

de **T**, o preditivista afirma:

(1) **P(T | E & ¬O) > P(T | E & O)**

Maher defende que a predição bem-sucedida dos 99 arremessos iniciais constitui evidência persuasiva de que o preditor 'tem um método confiável' de fazer predições de resultados de lançamento de moeda. A consistência de **T** com **E** no caso do acomodador não fornece nenhuma evidência particular de que o método do acomodador é confiável, assim, não temos nenhuma razão particular para endossar sua predição sobre a 100ª jogada. Permitindo que **R** afirme que o método em questão é confiável, e $\mathbf{M}_T$ que o método **M** gerou a hipótese **T**, isso leva a:

(2) **P(R | M_T & E & ¬O) > P(R | M_T & E & O)**

Maher (1988) fornece uma prova rigorosa de (2), que é vista implicando (1) em várias suposições.

Maher (1988) faz a suposição simplificada de que qualquer método de predição usado por um preditor é ou completamente confiável (esta é alegação abreviada por '**R**') ou não é melhor do que um método aleatório (**¬R**). (Maher [1990] mostra que essa suposição pode ser derrotada e graus contínuos de confiabilidade assumidos; ainda assim o resultado preditivista é gerado). Em termos qualitativos, onde **M** gera **T** (e, assim, prediz **E**) sem inserção da evidência **E**, devemos inferir que é muito mais provável que o método que gerou **E** seja confiável no lugar de **E** simplesmente ter calhado de ser verdadeiro apesar de que **R** não era melhor do que um método aleatório. Em outras palavras, julgamos ser muito mais provável nos depararmos com um sujeito usando um método confiável de predição de lançamento de moeda do que de nos depararmos com uma sequência de 99 predições verdadeiras de lançamento de moeda que eram meramente palpites de sorte, porque

$$\mathbf{P(R) \gg (1/2)^{99}(1-P(R))}.$$

Maher articulou um preditivismo heurístico fraco porque ele alega que o sucesso preditivo é sintomático do uso de um método confiável de descoberta.[39]

---

[39] Maher esclarece que a noção de novidade por ele buscada é heurística. (1993, p. 329).

Para uma discussão crítica da teoria do preditivismo de Maher, confira Howson e Frankling (1991) e a resposta de Maher (1993), Barnes (1996a,b), Lange (2001), Harker (2006) e Worrall (2014).[40]

6.2 A Explicação Forjada

Notou-se acima que hipóteses *ad hoc* permanecem sob suspeita por várias razões, uma das quais era a de que a hipótese que fosse proposta para resolver uma dificuldade particular pode não concordar bem com a teoria que se propôs a salvar ou com as crenças de fundo relevantes.[41] Isso poderia resultar do fato de que não há caminho óbvio para resolver a dificuldade de um modo que seja totalmente 'natural' de um ponto de vista da teoria mesma ou de critérios operativos de escolha de teoria. Por exemplo, a teoria do flogisto afirmava que substâncias emitiam flogisto enquanto queimavam. Contudo, foi estabelecido que algumas substâncias realmente ganhavam peso ao queimar. Para acomodar o último fenômeno, foi proposto que o flogisto tinha peso negativo, mas a última hipótese era claramente *ad hoc*, no sentido de falhar em concordar com a crença de fundo de que substâncias simplesmente não possuem peso negativo, e com o conhecimento de que muitos objetos perdem peso quando queimados (PARTINGTON; MCKIE, 1938a, p. 33-38).

Portanto, a "explicação forjada" defende o preditivismo, ao apontar que o processo de acomodação se presta à proposta de hipóteses que não concordam naturalmente com as restrições operativas na escolha de teorias, enquanto predições bem-sucedidas são imunes a essa preocupação (LIPTON 1990, 1991, cap.8). É claro que é uma questão importante saber se cientistas realmente confiam no fato de que a evidência foi predita (ou acomodada) em suas avaliações de teorias, se uma teoria foi forjada para acomodar algum dado, o cientista não poderia simplesmente notar que a teoria forjada sofre um defeito de coerência e não prestar atenção sobre

---

[40] Uma teoria de preditivismo de algum modo similar é fornecida por Kahn, Landsberg e Sockman (1990) a qual defende que o sucesso preditivo confirma que um cientista é 'talentoso' e, assim, é mais provável que endosse uma boa teoria.

[41] Lembra, portanto, *ad hoc*$_3$, de Zahar, que "é obtida de sua predecessora por uma modificação das hipóteses auxiliares que não concordam com o espírito da heurística do programa (1973, p. 101).

se os dados foram acomodados ou preditos? Alguns defendem que, embora aqueles cientistas sejam juízes imperfeitos de tal coerência, um cientista que acomoda algum dado pode pensar que sua acomodação seja totalmente coerente, enquanto seus pares podem ter uma visão mais precisa e objetiva de que ela não seja. O "apoio avaliado" que o cientista atribui à sua acomodação pode, portanto, falhar em coincidir com seu "apoio objetivo", e o cientista poder confiar no fato de que sua evidência fora acomodada como prova de que fora forjada, ou, ao contrário, de que sua evidência foi predita como prova de que não tenha sido forjada (LIPTON, 1991, p. 150).

6.3 Conjunções Arbitrárias e Não Arbitrárias

Lange (2001) oferece uma interpretação substituta do exemplo do lance da moeda que afirma que o processo de acomodação (diferente da predição) tende a gerar teorias que não são fortemente apoiadas pelos dados confirmantes. Ele imagina uma versão 'distorcida' do exemplo de lance de moeda no qual os 99 resultados iniciais formam uma sequência de alternância estrita "coroa, cara, coroa, cara...", ao invés de formar a "sequência aparentemente aleatória" de resultados fornecida no caso original. Novamente imaginamos um preditor que corretamente prediz antecipadamente 99 resultados e um acomodador que os testemunha. Tanto o preditor quanto o acomodador predizem que o 100º resultado será coroa. Agora há pouca ou nenhuma diferença entre nossa probabilidade avaliada de que o sujeito corretamente irá predizer o 100º resultado.

> Isso sugere que a diferença intuitiva entre o par de exemplos original de Maher não reflete uma diferença entre predição e acomodação por si (LANGE, 2001, p. 580).

A análise de Lange apela para o que Goodman chamou uma 'conjunção arbitrária' - cuja marca é que

> o estabelecimento de uma componente dota a afirmação toda com a não-credibilidade que é transmitida às outras proposições componentes (1983, p. 68-9).

Um exemplo de conjunção arbitrária é “O sol é feito de hélio e 3 de agosto de 2017 cai numa terça-feira e 17 é um número primo”. No caso original do lance de moeda, julgamos que **H** é fracamente apoiada no cenário do acomodador porque julgamos que a sequência aparentemente aleatória de resultados é provavelmente uma conjunção arbitrária - portanto, o fato de que os 99 conjuntos iniciais são confirmados implica quase nada sobre o que será o 100º resultado. Mas o sucesso do preditor em predizer os 99 resultados iniciais implica fortemente que a sequência não é uma conjunção arbitrária, afinal:

> agora acreditamos com maior probabilidade que o agente foi levado a supor sua sequência particular por meio de algo que não notamos que ligasse a sequência em conjunto – algo que seria resguardado de ser uma coincidência tal que a hipótese é precisa no 100º arremesso (LANGE, 2001, p. 581).

Ao julgar que essa não é uma conjunção arbitrária, estamos agora preparados para reconhecer os primeiros 99 resultados a confirmar fortemente a predição do 100º caso. O que conta para a diferença entre os dois cenários, em outras palavras, não é primariamente se **E** foi predita ou acomodada, mas se julgamos que **H** seja uma conjunção arbitrária, e dessa forma se **E** fornece apoio para a porção remanescente de **H**.

Portanto, no caso distorcido de Lange, a não existência do efeito preditivista é devido ao fato de que é claro a partir das 99 jogadas iniciais que a sequência não é uma conjunção arbitrária, portanto, **E** confirma **H** de modo igualmente forte em ambos os cenários.

Lange continua a sugerir que na ciência real a prática de construir uma hipótese por meio da acomodação da evidência conhecida tem uma tendência a gerar conjunções arbitrárias. Portanto, a hipótese da contração de Lorentz, quando anexada à sua eletrodinâmica para acomodar a falha em detectar opticamente qualquer movimento em relação ao éter, resultou em uma conjunção arbitrária (uma vez que a evidência que apoiava a hipótese da contração não apoiava a eletrodinâmica, ou vice-versa), essencialmente por essa razão, argumenta Lange, ela foi rejeitada por Einstein como *ad hoc*. Quando a evidência é predita por uma teoria, por contraste, é tipicamente porque a teoria não é uma conjunção arbitrária. A importância evidencial

da predição e da acomodação para Lange é que ambas tendem a estar correlacionadas (negativamente e positivamente) com a construção de teorias que são conjunções arbitrárias. A visão de Lange deve, portanto, ser classificada como um preditivismo heurístico fraco, apesar de Lange nunca tomar partido sobre se cientistas realmente confiam em tais correlações ao avaliar teorias.

Para discussões críticas da teoria de Lange, confira Worrall (2014, p. 59-61) e Harker (2006, p. 317).

6.4 Testes Severos

Deborah Mayo (1991, 1996 e 2014) defendeu que a intuição de que o preditivismo é verdadeiro deriva de um prêmio sobre testes severos de hipóteses. Um teste de uma hipótese **H** é severo na medida em que é improvável que **H** passe no teste, caso **H** seja falsa. Intuitivamente, se é mostrado que uma consequência nova **N** segue de **H**, e é muito baixa a probabilidade de **N** sobre a suposição de **~H** (pela razão de ser nova), então testar por **N** pareceria contar como um teste severo de **H**, e um resultado positivo deveria apoiar fortemente **H**. Aqui novidade e severidade parecem coincidir, mas Mayo observa que há casos nos quais elas estão separadas. Por exemplo, pareceu a muitos que se **H** é construída para encaixar algum corpo de evidência **E**, então o fato de que **H** se ajusta a **E** não apoia **H** porque esse ajuste não constitui que **H** tenha sobrevivido a um teste severo (ou a um teste em geral). Um dos objetivos centrais de Mayo é expor as falácias que este último raciocínio envolve.

Giere (1984, p. 161-163) afirma que a evidência para cujo encaixe **H** foi construída não apoia **H** porque, dado o modo como **H** foi construída, está destinada a encaixar aquela evidência. Mayo resume seu raciocínio como se segue:

> (1) Se **H** é construída para o uso [*use-constructed*], então, um encaixe bem-sucedido é assegurado não importa qual.

Mas Mayo nota que "não importa qual" pode ser interpretado de dois modos: (a) não importa quais são os dados, e (b) não importa se **H** é verdadeira ou falsa. (1) é verdadeira quando interpretada como (a), mas para estabelecer que a evidência acomodada falha em apoiar **H** (como Giere pretende) (1) precisa ser interpretado como (b). Entretanto, (1) é falsa quando interpretada assim. Mayo (1996, p. 271)

ilustra isso com um exemplo simples: seja a evidência e uma lista de notas do SAT[42] , tiradas pelos estudantes de uma turma em particular. Usar essa evidência para computar a nota média *x* e configurar **h = a** pontuação média no SAT para esses estudantes é **x**. Agora é claro que **h** foi construída para o uso [*use-constructed*] a partir de **e**. É verdadeiro que qualquer pontuação média que fosse computada encaixaria os dados, não importa que dados são esses, mas dificilmente seria verdadeiro que **h** encaixaria os dados não importa se **h** fosse verdadeira ou falsa. Se **h** fosse falsa não encaixaria os dados, porque os dados iriam, inevitavelmente, ajustar-se apenas a uma hipótese verdadeira. Portanto, **h** passou em um teste maximamente severo: é virtualmente impossível para **h** encaixar os dados se **h** for falsa, apesar do fato de **h** ser feita para encaixar **e**.

Mayo dá um exemplo adicional de como uma hipótese, construída para o uso [*use-constructed*], pode contar como tendo sobrevivido a um teste severo pertencente ao famoso experimento da *Teoria Geral da Relatividade*, de Einstein, no eclipse de 1919. A TGR predisse que a luz das estrelas que passasse pelo sol se curvaria a um grau determinado (especificamente 1,75 arco-segundos). Houve realmente duas expedições realizadas durante o eclipse, uma a Sobral, nordeste[43] do Brasil e a outra à ilha do Príncipe, no Golfo da Guiné. Cada expedição gerou um resultado que apoiou a TGR, mas houve um terceiro resultado gerado pela expedição de Sobral que parecia refutar a TGR. Esse resultado foi, no entanto, desqualificado porque foi determinado que um espelho usado para captar imagens das posições das estrelas fora danificado pelo calor do sol. Embora alguém possa se preocupar que tal rejeição de evidências anômalas foi o tipo de ajuste *ad hoc* contra o qual Popper advertiu, Mayo nota que este é um caso perfeitamente legítimo de uso da evidência para apoiar uma hipótese (de que o terceiro resultado não era confiável) que equivale a dizer que a hipótese passou em um teste severo. Mayo conclui que uma proibição geral sobre hipóteses construídas para o uso "falham em distinguir entre construções para o uso problemáticas e não problemáticas (ou de dupla contagem)" (1996, p. 285). Entretanto, Hudson (2003) defende a existência de evidências históricas, sugerindo haver razões legítimas para questionar a hipótese

---

[42] N.T: Trata-se de um exame de cunho nacional, realizado nos EUA, para estudantes do equivalente brasileiro ao Ensino Médio.

[43] N.T.: O autor diz *Northern Brazil*, isto é, norte do Brasil. Mas está claramente equivocado, uma vez que Sobral fica no estado do Ceará, pertencente à região Nordeste.

de que o terceiro resultado não era confiável (ele usa esse ponto em apoio à sua própria alegação de que o fato de uma hipótese ser construída para o uso é evidência prima facie de que a hipótese é suspeita). Mayo (2003) responde que na medida em que o terceiro resultado era, não obstante, suspeito, os físicos envolvidos estavam certos em descartá-lo.

Mayo (1996, cap. 9) defende uma posição similar ao preditivismo atribuída aos métodos estatísticos de Neyman-Pearson, a proibição sobre construções de hipóteses pós-teste. Para ilustrar, Kish (1959) descreve um estudo que investigou a relação estatística entre um grande número de experiências de treinamento infantil (enfermagem, treinamento de banheiro, desmame, etc.) e os traços subsequentes de personalidade e comportamento (por ex., o ajustamento à escola, o roer de unhas, etc.). O estudo encontrou um número de altas correlações entre certas experiências de treinamento e os traços posteriores. O problema era que o estudo investigara tantas experiências de treinamento que era bastante provável que algumas correlações apareceriam nos dados simplesmente por acaso, mesmo se tivesse sido provado definitivamente não haver tal correlação. Um investigador que estudasse muitas possíveis correlações poderia, portanto, coletar aqueles dados e simplesmente olhar para as diferenças estatisticamente significantes e proclamar evidência para correlações, apesar de tal evidência ser enganadora, envolvendo-se, portanto, na prática dúbia de 'construção de hipótese pós-teste'[44] . Mayo nota que tais hipóteses não deveriam contar como tendo passado em testes severos, endossando, portanto, a proibição de Neyman-Pearson acerca de tais construções. Hitchcock e Sober (2004) notam que a definição de severidade de Mayo, como aplicada nesse caso, difere-se da que ela emprega ao lidar com casos similares ao seu exemplo do SAT; Mayo (2008) responde à altura de seus críticos e argumenta que ela emprega duas versões de definição de severidade, enquanto eles, não obstante, refletem uma concepção unificada de severidade.

Para uma discussão crítica da explicação de Mayo, confira Iseda (1999) e Worrall (2006, p. 56-60, 2010, p. 145-153). Consulte, também, as respostas de Mayo a Worrall (1996, p. 265, 2010).

---

[44] Em uma veia de algum modo similar, Barnes (1996b, 2008, cap.5) estuda a diminuição de impacto do sucesso preditivo quando grandes quantidades de predições são feitas.

6.5 Confirmação Condicional e Incondicional

John Worrall tem sido um contribuidor importante para a literatura preditivista desde os anos 1970 até o tempo presente. Ele, ao lado de Elie Zahar, foi um dos primeiros proponentes da importância heurística da novidade (WORRALL, 1978, 1985). Em seus trabalhos mais recentes (1989, 2002, 2005, 2006, 2010, 2014) e, também, Scerri e Worrall (2001), Worrall estabeleceu uma detalhada teoria do preditivismo que, embora algumas vezes apresentado em termos heurísticos, é "na origem uma teoria lógica da confirmação" (2005, p. 819), é, portanto, uma explicação heurística fraca que toma a novidade-uso da evidência como sintomática de características lógicas ressaltadas que estabelecem a confirmação forte da teoria.

A explicação madura de Worrall é baseada em uma visão sobre teorias científicas que ele atribui a Duhem, o qual alega que uma teoria científica seja naturalmente pensada como consistindo em um enunciado central em união com certo conjunto de enunciados auxiliares mais específicos. É comumente o caso em que a teoria central deixará indeterminados certos "parâmetros livres" e os enunciados auxiliares fixem valores para tais parâmetros. Para citar um exemplo muito usado por Worrall, a teoria ondulatória da luz consiste na teoria central de que a luz é uma perturbação periódica transmitida através de algum tipo de meio elástico. Esse enunciado central deixa, por si só, em aberto vários parâmetros livres a respeito dos comprimentos de onda de tipos particulares de luz monocromática. Worrall propõe a compreensão do status diminuto do apoio evidencial associado com a acomodação como se segue: quando a evidência e é "usada" na construção de uma teoria, é tipicamente utilizada para estabelecer o valor de um parâmetro livre em alguma teoria central **T**. A versão corrigida será uma versão **T'** específica de **T**. e serve para confirmar **T'**, então, somente na condição de que haja apoio independente para **T**, portanto, a acomodação fornece apenas "confirmação condicional". De modo relevante, a evidência e que é usada dessa forma não fornecerá, por si mesma, evidência para a teoria central **T**. Worrall (2002, p. 201) oferece como uma ilustração o apoio dado à teoria ondulatória da luz (**W**) pelo experimento da dupla fenda utilizando a luz de um arco de sódio - os dados consistirão em várias alternâncias entre luz e 'franjas' escuras. Os dados da franja podem ser usados para calcular o comprimento de onda da luz de sódio, e, portanto, utilizados para gerar uma versão mais específica da teoria ondulatória da luz **W'**, uma que reúna **W** a um enunciado

sobre o comprimento de onda desse tipo específico de luz. Mas os dados dão meramente apoio condicional para **W'**, ou seja, os dados apoiam **W'** apenas na condição de que haja evidência independente para **W**.

A evidência prevista, para Worrall, é, portanto, a evidência que não é utilizada para fixar os parâmetros livres. Worrall cita duas formas que as predições podem tomar: uma é quando uma consequência evidencial particular decorre "imediatamente da parte central", isto é, é uma consequência da parte central, em conjunção com "auxiliares naturais", e a outra é quando ela é uma consequência de uma versão específica de uma teoria cujos parâmetros foram fixados pelo uso de outros dados. Para ilustrar a primeira: o movimento retrógrado[45] era uma consequência natural do núcleo copernicano (a afirmação de que a terra e os planetas orbitam em torno do sol) porque a observação dos planetas era realizada sobre um observatório móvel que periodicamente passava por outros planetas, embora isso pudesse apenas ser acomodado pela astronomia ptolomaica pela sugestão e o ajuste de hipóteses auxiliares que supusessem que os planetas se movem em epiciclos (o movimento retrógrado não se segue naturalmente da ideia central ptolomaica de que o Sol, estrelas e planetas orbitam a terra). Portanto, o movimento retrógrado foi predito pela teoria copernicana e, assim, deu apoio incondicional àquela teoria, ao passo que deu apenas confirmação condicional à teoria ptolomaica. A segunda forma de predição é aquela que se segue de uma versão específica de uma teoria que não fora, entretanto, utilizada para fixar um parâmetro, imagine que **W'** do parágrafo anterior faça uma nova predição **p** (digamos para outro experimento, tal qual o experimento de uma fenda), **p** oferece confirmação incondicional a **W'** e a **W** (WORRALL, 2002, p. 203).

No entanto, é importante entender que a repetida expressão de Worrall sobre sua posição ser uma concepção heurística de novidade (particularmente depois de seu texto de 1985) não equivale a um endosso do preditivismo heurístico forte. Worrall esclarece isso em seu artigo de 1989 que enfoca a importância evidencial da 'mancha branca' na confirmação da versão da teoria ondulatória da luz de Fresnel. A razão pela qual o dado da mancha branca carregou tanto peso

---

[45] "Movimento retrógrado" se refere à relativa posição observada dos planetas contra as estrelas fixas, um movimento planetário (observado ao longo de uma sequência de noites) relativo às estrelas cessará periodicamente, reverterá a direção, e então retoma sua direção original.

não é, definitivamente, a de que Fresnel não o tenha utilizado na construção da teoria, mas porque esse dado se segue naturalmente da teoria central de que a luz é uma onda. A razão pela qual os dados da franja usados para calcular o comprimento de onda da luz de sódio (confira acima) não terem tanto peso é que não eram uma consequência dessa ideia central (nem o comprimento de onda da luz de sódio fora fixado por quaisquer outros dados). Portanto, d é novo para **T** quando "há uma via heurística para [**T**] a qual não pressupõe a existência [de **d**]." (SCERRI; WORRALL, 2001, p. 418). Como Worral, às vezes, coloca, saber se **d** carrega confirmação incondicional para **T** não depende de se **d** foi realmente usada na construção de **T**, mas se foi "necessária" para construir **T** (1989, p. 149-151). Assim Worrall é definitivamente um proponente do preditivismo heurístico fraco (não forte). Worrall rejeita categoricamente o preditivismo temporal, argumentando que o fato de que a mancha branca fora uma consequência temporalmente nova não era, por si só, de importância epistêmica.

Para discussões adicionais da teoria do preditivismo de Worrall, confira Mayo (2010, p. 155), Schurz (2014), Votsis (2014) e Douglas e Magnus (2013, p. 587-588).

Scerri e Worrall (2001) apresentam uma renderização detalhada do episódio histórico da avaliação da teoria da lei periódica de Mendeleev pela comunidade científica, no qual foi defendido que esse caso justifica completamente a teoria do preditivismo de Worrall.

Para discussão de Scerri e Worrall, consulte Akeroyd (2003), Barnes (2005b), e as respostas de Worrall (2005) e Scerri (2005), além dos trabalhos de Schindler (2008, 2014) e Brush (2007).

6.6 A Analogia do Arqueiro

Um argumento comum a favor do preditivismo é o de que devemos evitar inferir que uma teoria **T** é verdadeira com base na evidência **E** para cujo encaixe ela foi feita, uma vez que podemos explicar o porquê de **T** implicar **E** simplesmente notando como **T** foi construída - mas se **T** não foi construída para acomodar **E** então só a verdade de **T** pode explicar o fato de que **T** se adequa a **E**. Vários filósofos notaram que esse raciocínio é falacioso. Como visto acima, não faz sentido oferecer uma explicação (por exemplo, em termos de como a teoria foi construída) pelo fato de **T** implicar **E** - pois esse último fato é um fato lógico para o qual nenhuma explicação causal pode ser dada. Na medida em que há aqui um *explanandum* em

busca de um *explanans*, trata-se muito mais do fato de que o teórico conseguiu construir ou "escolher" uma teoria (que veio a ser **T**), a qual implicava corretamente **E** (COLLINS, 1994; BARNES, 2002), aquele *explanandum* poderia ser explicado pela observação de que o teórico construiu uma teoria (que veio a ser **T**) para encaixar **E**, ou endossou-a porque ela encaixou E.

White (2003) oferece uma teoria do preditivistmo que começa com o mesmo *insight* - o *explanandum* relevante é:

> (ES) O teórico selecionou uma teoria que implica o dado.
>
> Esse *explanandum* pode ser explicado de duas maneiras:
>
> (DS) O teórico projetou sua teoria para implicar o dado.
> (RA) A seleção da teoria pelo teórico visava de forma confiável a verdade.

White explica que (RA) significa "grosso modo, que os mecanismos que levaram à sua seleção de uma teoria deram-lhe uma boa chance de chegar à verdade" (2003, p. 664). Assim, White faz uma analogia entre o teórico e um "arqueiro" que é mais ou menos confiável em **mirar** na verdade, ao selecionar uma teoria. Então, White oferece um simples argumento em favor do preditivismo: assumindo ~DS, ES fornece evidência para RA. Mas assumindo DS, ES não fornece nenhuma evidência para RA. Portanto, o preditivismo heurístico é verdadeiro.

Curiosamente, White nomeia sua descrição como uma descrição heurística forte. Ao fazer essa afirmação ele está afirmando que a vantagem epistêmica da predição não seria totalmente apagada para um observador que estivesse completamente ciente de toda evidência relevante e do conhecimento de fundo possuído pela comunidade científica no período relevante do tempo. Isso porque o grau de acordo com o qual a teorização é confiável depende de princípios de avaliação da evidência e de relações causais (incluindo a confiabilidade de nossas faculdades perceptuais, a precisão dos instrumentos de medida, etc.) que não são totalmente "transparentes" para nós[46] . Na

---

[46] Barnes (2008, cap. 2) defende uma afirmação similar. Compare, também, a distinção de Lipton entre "apoio avaliado" e "apoio real" (*vide* Seção 6.2).

medida em que cientistas totalmente informados podem não estar totalmente convencidos sobre quão confiáveis são esses princípios e relações, a evidência de que eles são levados a endossar teorias que sejam preditivamente bem-sucedidas continua a redundar em sua avaliação de confiabilidade. Portanto, White conclui, o preditivismo heurístico forte é justificado (2003, p. 671-674).

6.7 A Abordagem de Akaike

Hitchcock e Sober (2004) fornecem uma teoria original do preditivismo heurístico fraco que é baseado em uma preocupação particular com a acomodação. Na suposição de que os dados são ruidosos, ou seja, imbuídos com erros observacionais, uma boa teoria nunca irá encaixar os dados perfeitamente. Construir uma teoria que encaixe os dados melhor do que deveria fazer uma boa teoria, considerando os dados ruidosos, é ser culpado de "encaixe exagerado" [*overfitting*], se sabemos que uma teórica construiu sua teoria para acomodar os dados, bem podemos nos preocupar sobre ela ter encaixado exageradamente [*overfit*] os dados e, desse modo, construído uma teoria falha. Se sabemos, entretanto, que uma teórica construiu sua teoria sem acesso a tais dados, ou sem utilizá-los no processo de construção da teoria, não precisamos nos preocupar que tenha ocorrido encaixe exagerado dos dados. Quando tal teoria continua a fazer predições bem-sucedidas, Hitchcock e Sober continuam a defender, isso nos dá evidência de que os dados nos quais a teoria estava inicialmente baseada não foram exageradamente encaixados no processo de construção da teoria.

A abordagem de Hitchcock e Sober deriva de uma solução particular para o problema do ajuste da curva apresentado em Foster e Sober (1994). O problema do ajuste da curva é saber como selecionar uma curva otimamente apoiada com base em um corpo de dados qualquer (por exemplo, um conjunto de pontos [**X**, **Y**] traçados num gráfico de coordenadas). Uma curva bem-apoiada reunirá tanto "ajustabilidade" [*goodness of fit*] aos dados e simplicidade (intuitivamente, evitando altas protuberâncias ou padrões irregulares). Resolver o problema do ajuste da curva requer alguma maneira precisa de caracterizar a simplicidade da curva, uma maneira de caracterizar sua ajustabilidade, e um método de balancear a simplicidade contra a ajustabilidade para identificar uma curva ótima.

Foster e Sober citam o resultado de Akaike (1973), de que uma estimativa imparcial da precisão preditiva de um modelo pode ser calculada pela avaliação

tanto de sua ajustabilidade quanto de sua simplicidade como medida pelo número de parâmetros ajustáveis que ela contém. Um modelo é um enunciado (um polinômio, no caso da curva proposta) que contém pelo menos um parâmetro ajustável. Para qualquer modelo **M** em particular, um conjunto de dados qualquer, e identificando **L***(***M***)* como a curva mais provável, ou seja, que melhor se ajusta aos dados de **M**, Akaike mostrou que:

> Uma estimativa imparcial da precisão preditiva do modelo
>
> **M ≈ log[Pr(Dados | L(M))] − k**.

Essa estimativa é considerada como uma contagem do 'Critério de Informação de Akaike' (CIA) para modelos, ela mede a ajustabilidade em termos do *log* da probabilidade dos dados na suposição de **L***(***M**). A simplicidade do modelo é inversamente proporcional a **k**, o número de parâmetros ajustáveis no modelo. A ideia intuitiva é que modelos com um alto valor de **k** fornecerão uma grande variedade de curvas que tenderão a encaixar os dados mais aproximadamente do que modelos com um valor de **k** mais baixo, e, assim, valores altos de **k** são mais propensos ao encaixe exagerado [overfitting] do que pequenos valores de **k**. Então, a contagem do CIA avalia a probabilidade de precisão preditiva de um modelo de uma maneira que equilibra tanto a ajustabilidade quanto a simplicidade, e o problema do ajuste da curva é indiscutivelmente resolvido.

Hitchcock e Sober (2004) consideram um exemplo hipotético envolvendo duas cientistas, Penny Preditora e Annie Acomodadora. Trabalhando independentemente, elas recebem o mesmo conjunto de dados **D**, Penny propõe a teoria $\mathbf{T_p}$ enquanto Annie propõe $\mathbf{T_a}$. A diferença crítica, entretanto, foi que Penny propôs $\mathbf{T_p}$ com base em um segmento inicial dos dados **D1**, depois disso ela predisse os dados restantes **D2** com alto grau de precisão (**D** = **D1** ∪ **D2**). Annie, no entanto, tinha em mãos todos os dados **D** antes de propor $\mathbf{T_a}$ e, ao propor essa teoria acomodou **D**. Hitchcock e Sober perguntam se deve haver razão para suspeitar que a teoria de Penny será mais preditivamente precisa no futuro e, nesse sentido exato, ser mais bem-confirmada.

Hitchcock e Sober defendem que não há resposta para essa questão – e, então, apresentam uma série de vários casos. Na medida em que o preditivismo se sustenta em alguns e não em outros, sua descrição do preditivismo é claramente uma descrição local (ao invés de global). Em casos nos quais Penny e Annie propõem

a mesma teoria, ou propõem teorias cujas contagens de CIA podem ser calculadas e comparadas diretamente, não há razão para considerar que fatos sobre como elas construíram suas teorias tenham importância mais adiante. Mas se não sabemos quais teorias foram propostas, ou por quais métodos foram construídas, o fato de que Penny predisse os dados que Annie acomodou pode pender em favor de a teoria de Penny possuir uma contagem de CIA maior do que a de Annie e, dessa forma, levar uma vantagem epistêmica.

Na medida em que o preditivismo se sustenta em alguns casos mas não nos outros, a questão sobre se o preditivismo está ancorado em episódios reais da ciência depende de quais casos tais episódios reais tendem a reunir, mas Hitchcock e Sober "não assumem qualquer lado sobre quão frequentemente os vários casos emergem" (2004, p. 21).

Apesar de sua descrição de preditivismo ser inicialmente feita sob medida para o problema do ajuste da curva, não é de modo algum restrito a tais casos. Eles notam que é natural pensar que um modelo seja análogo ao quadro ontológico de uma teoria científica onde os vários comprometimentos ontológicos podem funcionar como "parâmetros ajustáveis", por exemplo, os retratos de mundo ptolomaico e copernicano começam, ambos, com a afirmação de que uma certa entidade (o sol ou a terra) está no centro, e esses modelos são articulados ao produzir modelos com parâmetros ajustáveis.

Para discussão crítica da explicação de Sober e Hitchcock, confira Lee (2012, 2013) e Douglas e Magnus (2013, p. 582-584).

6.8. A Novidade de Endosso e a Confirmação de Crenças de Fundo

Barnes (2005a, 2008) sustenta que o preditivismo é frequentemente a manifestação de um fenômeno que ele intitula "pluralismo epistêmico". Uma 'avaliadora de **T'** (uma cientista que atribui alguma probabilidade à teoria **T**) é uma pluralista epistêmica na medida em que considera que uma forma de evidência sejam as probabilidades publicadas, isto é, publicamente apresentadas, por outros cientistas a favor e contra **T** e outras afirmações relevantes (ela é uma individualista epistêmica se não faz isso, mas apenas considera a evidência científica "por si mesma"). Uma forma de evidência pluralística é o evento no qual um cientista respeitável endossa uma teoria - isso ocorre quando um cientista publica uma probabilidade a favor de **T** que (1) não seja menor do que a probabilidade do

avaliador e (2) seja alta o suficiente para que a confirmação preditiva subsequente de **T** redunde na credibilidade do cientista. (2008, p. 2.2)

Barnes rejeita a concepção heurística de novidade com o fundamento de que é um erro pensar que o que importa epistemicamente seja o processo pelo qual a teoria fora construída, o que importa é em que base a teoria fora endossada (2008, p. 33). No exemplo acima, a confirmação de **N** (uma consequência de **T**) pôde carregar um peso especial para um avaliador que compreendeu que o teórico endossou a teoria sem apelo à evidência observacional em favor de **N** (não, respectivamente, a como a teoria fora construída). Ele propõe a substituição da concepção heurística pela sua concepção de novidade de endosso: **N** (uma conhecida consequência de **T**) conta como uma confirmação nova de **T** em relação ao agente **X** na medida em que **X** publica uma probabilidade de nível de endosso para **T** que seja baseada num corpo de evidência que não inclua evidência baseada na observação em favor de **N**.

Barnes afirma que a noção de novidade de endosso possui algumas vantagens sobre a concepção heurística - uma delas é que a novidade de endosso pode contar a favor do fato de que a predição é uma questão de grau: quanto mais fortemente um teórico endossa **T**, mais fortemente sua consequência **N** é predita (e, portanto, mais evidência favorável a **T** para os avaliadores pluralistas que confiam no endossador). Outra vantagem é que a distinção ortodoxa entre contexto da descoberta e contexto da justificação é preservada. De acordo com a última distinção, para fins de avaliação de teoria não importa como uma teoria foi descoberta. Mas isso não se mostra verdadeiro na concepção heurística, dada a importância central acordada por ela para o modo como uma teoria foi construída (LEPLIN, 1987). A novidade de endosso respeita a irrelevância do processo pelo qual teorias são descobertas (BARNES, 2008, p. 37-38).

Uma afirmação central para essa descrição é a de que a confirmação é uma relação tripartida entre teoria, evidência e crença de fundo (GOOD, 1967). Barnes diferencia dois tipos de endossadores de teorias: (1) endossadores virtuosos que publicam probabilidades para as teorias coerentes com suas evidências e crenças de fundo e (2) endossadores não virtuosos que publicam probabilidades não tão coerentes. Uma maneira comum de explicar a intuição preditivista é notar que acomodadores tendem a ser vistos com certa suspeita, seu endosso de **T** baseada em evidência acomodada pode refletir um tipo de pressão social para endossar **T** qualquer que seja seu mérito (confira a "explicação forjada" acima). Tal endossador pode publicar uma probabilidade em favor de **T** que seja alta demais,

dada sua evidência total e sua crença de fundo - o preditivismo, assim, se torna uma estratégia pela qual endossadores pluralistas se protegem de acomodadores não virtuosos (BARNES, 2008, p. 61-69).

Barnes, então, apresenta uma teoria do preditivismo que é planejada para se aplicar a endossadores virtuosos. O preditivismo virtuoso têm duas origens: (1) a predição por si mesma, a qual é constituída por um endossador publicando uma probabilidade de nível de endosso a favor de **T**, que implica a consequência empírica **N**, com uma fundamentação que não inclui evidência baseada em observação para **T***;* e (2) o sucesso preditivo, constituído pela demonstração empírica de que **N** é verdadeira. A predição por si mesma carrega importância epistêmica em favor do endossador pluralista porque ela implica que o preditor possui a razão **R** (consistindo em crenças de fundo) que apoiam **T**. Se o endossador vê o preditor como confiável, este simples ato de predição carrega peso epistêmico. O sucesso preditivo, então, confirma a verdade de **R**, que assim passa a contar como evidência favorável a **T**. A nova confirmação, portanto, tem a virtude especial de confirmar as crenças de fundo do preditor, falta à confirmação acomodativa tal virtude.

Barnes apresenta dois experimentos de pensamento bayesianos com a finalidade de estabelecer o preditivismo virtuoso. Em cada experimento uma avaliadora Eva se depara com dois cenários, um no qual ela confronta Peter, o qual publica uma probabilidade de endosso favorável a **T** sem apelar a observações em apoio a **N** (assim, Peter prediz **N**), e outro, no qual ela confronta Alex que publica uma probabilidade de endosso em favor de **T** numa base que inclui observações que estabelecem **N** (assim Alex acomoda **N**). A ideia por trás de ambos os experimentos de pensamento é tornar os cenários, de outra forma, tão similares quanto possível - Barnes faz uma série de suposições *ceteris paribus* (todo o resto igual) que dão às funções de probabilidade de Peter e Alex semelhança máxima. Contudo, mostrou-se que há mais de uma maneira de manter os cenários maximamente similares: no primeiro experimento, Peter e Alex possuem a mesma proporção de probabilidade relativa [*likelihood*], mas possuem diferentes probabilidades posteriores. No segundo cenário, eles possuem as mesmas probabilidades posteriores, mas diferentes proporções de probabilidade relativa. Barnes demonstra que a probabilidade posterior de Eva é maior no cenário preditor em ambos os experimentos, justificando assim o preditivismo virtuoso (2008, p. 69-80).

Embora sua defesa do preditivismo virtuoso seja a peça central de sua explicação, Barnes alega que o preditivismo pode se mostrar verdadeiro sobre a avaliação de teorias reais em uma variedade de formas. Ele sustenta que a posição

considerada “preditivismo fraco” é realmente ambígua, pois poderia se referir à alegação de que cientistas realmente confiam no conhecimento de que a evidência foi (ou não) predita porque a predição é sintomática de alguma(s) outra(s) característica(s) epistemicamente importante(s) (“preditivismo temperado”[47]) ou simplesmente pelo fato de que há uma correlação entre predição e essa(s) outra(s) característica(s) (“preditivismo fino”). A distinção entre preditivismo temperado e fino combina-se com a distinção entre preditivismo virtuoso e não virtuoso para produzir quatro variedades de preditivismo fraco. Barnes, então, se volta para o caso da lei periódica de Mendeleev e defende que todas as quatro variedades podem ser diferenciadas na reação da comunidade científica à teoria dos elementos de Mendeleev (2008, p. 82-122). Em particular, ele argumenta que foi especificamente a evidência predita por Mendeleev, e não sua evidência acomodada, que teve o poder de confirmar, pelo ponto de vista da comunidade científica, as suas crenças de fundo científicas e metodológicas.

Respostas críticas à descrição de Barnes são apresentadas em Glymour (2008), Leplin (2009) e Harker (2011). Barnes (2014) responde a elas. Confira, também, Magnus (2011) e Alai (2016).

7. Antipreditivismo

Foi observado, na seção 1, que John Maynard Keynes rejeitou o preditivismo, defendendo que, quando uma teoria **T** é inicialmente construída, é usual que haja razões **R** em favor de **T**. Se **T** continua a gerar novas predições bem-sucedidas **E**, então tais razões se combinam com **R** para apoiar **T**, mas se alguma **T’** for construída ‘meramente para encaixar **E’**, então **T’** será menos apoiada do que **T**. Essa foi considerada a “dissolução keynesiana do paradoxo do preditivismo” (BARNES, 2008, p. 15-18).

Colin Howson cita com aprovação a dissolução keynesiana (1988: 382) e fornece a seguinte ilustração: considere **h** e **h’** como quadros explanatórios rivais. **h’** prediz e independentemente; **h** não implica e, mas possui um parâmetro livre que é fixado com base em **e** para produzir $\mathbf{h}(\mathbf{a}_0)$, essa última hipótese, então, implica

[47] Isso é o que Schindler (2014, p. 67) chama de “preditivismo sintomático”.

e. Assim, **h'** prediz **e** enquanto **h**(**a**$_0$) meramente acomoda **e**. Assumamos que as probabilidades prévias de **h** e **h'** sejam iguais, ou seja, **p**(**h**) = **p**(**h'**). Agora é lógico que **p**(**h**(**a**$_0$)) < **p**(**h**), uma vez que **h**(**a**$_0$) implica **h**, mas não o inverso. Assim, Howson mostra se seguir daí que o efeito da confirmação de **e** não deixará **h'** menos provável, e, muito possivelmente mais provável, que **h**(**a**$_0$) (1990, p. 236-237). Assim, o preditivismo parece verdadeiro, mas o fator operante é o papel desigual das probabilidades prévias.[48]

O argumento de Keynes e Howson contra o preditivismo sustenta que a evidência que parece apoiar o preditivismo é ilusória, eles claramente estão afirmando que o preditivismo forte é falso, presumivelmente em suas formas temporal e heurística.

No entanto, é importante notar que os argumentos de Keynes e Howson citados acima antecipam a introdução do conceito de "preditivismo fraco" na literatura.[49] Não é, assim, claro que posição Keynes e Howson tomariam sobre o preditivismo fraco. Da mesma forma, o artigo de Collins (1994) *Against the Epistemic Value of Prediction* [*Contra o valor epistêmico da predição*] rejeita fortemente o preditivismo, mas o que ele claramente nega é aquilo que foi desde então considerado preditivismo heurístico forte. Ele deve endossar o preditivismo heurístico fraco, já que concede que

> todos os lados do debate concordam que saber que uma teoria predisse um conjunto de dados, no lugar de tê-los acomodado, nos dá uma razão a mais para acreditar que ela seja verdadeira, informando-nos algo sobre as características estruturais/relacionais de uma teoria (1994, p. 213).

---

[48] Howson (1984, p. 249) oferece um exemplo similar igualando **h** com a teoria gravitacional newtoniana, **h**' com a relatividade geral, e uma descrição da órbita solar de Mercúrio, e **h**(**a**$_0$) com uma versão ajustada da teoria gravitacional newtoniana que inclui um postulado sobre a elipticidade solar que é fixada para explicar **e**. A alegação é que **h'** prediz e independentemente enquanto **h** meramente a acomoda, o resultado é **h** ser mais bem-confirmada por **e** do que **h**(**a**$_0$).

[49] Da mesma forma, Schlesinger (1987) rejeita redondamente o preditivismo temporal e o heurístico, mas parece claro que são as versões fortes de preditivismo que ele está rejeitando.

Harker similarmente defende que “é hora de deixar o preditivismo para trás” mas também concede que “algumas teses preditivistas fracas podem ser corretas”(2008, p. 451); Harker se preocupa com que a proclamação do preditivismo fraco possa confundir alguns fazendo-os pensar que o sucesso preditivo seja de algum modo mais importante do que outros indicadores epistêmicos (tais como o endosso por cientistas confiáveis). White vai tão longe quanto ao afirmar que o preditivismo fraco “não é controversial” (2003, p. 656).

Stephen Brush é o autor de um corpulento trabalho histórico, muito do qual visando mostrar que o preditivismo temporal não se sustenta em vários episódios da história da ciência.[50] Estes incluem o caso da curvatura da luz das estrelas na avaliação da *Teoria da Relatividade Geral* (BRUSH, 1989), as teorias dos fenômenos do plasma espacial de Alfvén, e o ressurgimento da cosmologia do *big bang* (BRUSH, 1993). Entretanto, Brush (1996) argumenta que a novidade temporal desempenhou um papel na aceitação da Tabela Periódica de Mendeleev baseada nas predições de Mendeleev. Scerri e Worrall (2001) apresentam considerável detalhe histórico sobre a avaliação da teoria de Mendeleev e contestam a afirmação de Brush de que novidade temporal tenha desempenhado importante papel na aceitação da teoria (2001, p. 428-436). Confira, também, Brush (2007).

8. O Debate Realista/Antirrealista

O realismo científico sustenta que haja evidência suficiente para acreditar que as teorias das “ciências maduras” são ao menos aproximadamente verdadeiras. Apelos para a novidade foram importantes na formulação de dois argumentos em favor do realismo científico, estes são o “argumento sem-milagre” e a resposta realista à chamada “indução pessimista”[51] .

O argumento sem-milagre a favor do realismo científico alega que o realismo é a única explicação que não faz do sucesso da ciência um milagre (PUTNAM,

---

[50] Para um resumo de algumas se suas pesquisas, confira Brush (1994).

[51] Muita literatura sobre este tópico discute a relevância do “sucesso preditivo” ou “sucesso empírico” para inferências à verdade da teoria sem estipular ou enfatizar um papel especial para os novos sucessos. Nesta seção o foco está na literatura que enfatiza um papel crítico para a novidade.

1975, p. 73). “O sucesso da ciência” aqui se refere à miríade de consequências empíricas das teorias maduras verificadas – mas, como vimos, há uma tendência de longa data em considerar com suspeição aquelas consequências empíricas verificadas que a teoria foi construída para encaixar. Assim, o “argumento definitivo a favor do realismo científico” refere-se a uma versão do argumento sem-milagre que se concentra apenas nas novas consequências das teorias, seria um milagre, proclama este argumento, uma teoria conseguir ter um registro sustentado de predições novas bem-sucedidas se a teoria não fosse ao menos aproximadamente verdadeira. Portanto, assumindo que não haja teorias competidoras com registros comparáveis de novos sucessos, devemos inferir que tais teorias são ao menos aproximadamente verdadeiras (MUSGRAVE, 1988)[52] .

Na medida em que o argumento definitivo para o realismo claramente enfatiza um papel especial para os novos sucessos, a natureza da novidade tem sido um foco importante na explicação realista. Leplin (1997) traz uma articulação, grande o bastante, do argumento definitivo a favor do realismo; Leplin propõe uma condição suficiente para a novidade consistindo em duas condições:

> Um resultado observacional **O** é novo para **T** se:
>
> Condição Independente: Há uma reconstrução minimamente adequada do argumento levando a **T** que não cita qualquer generalização qualitativa de **O**.
>
> Condição de Singularidade: Há alguma generalização qualitativa de **O** que **T** explica e prediz, e cujas instâncias nenhuma teoria alternativa fornece razão viável para esperar, no momento em que **T** inaugura tal feito (LEPLIN, 1997, p. 77).

---

[52] Foi observado que o argumento definitivo comete a falácia da taxa de base se ele apela ao sucesso empírico (incluindo o novo sucesso) como um argumento para a verdade da teoria na ausência de informação sobre a frequência com que teorias verdadeiras são propostas entre a totalidade de teorias propostas (Magnus e Calender 2004). Para discussões a respeito, confira Dicken (2013), Menke (2014), Henderson (2017) e Dawid e Hartman (2017).

Leplin esclarece que uma "reconstrução minimamente adequada" de tal raciocínio será uma dedução válida ***D*** das 'hipóteses identificadoras básicas' de **T** a partir de suposições de fundo garantidas de modo independente - as premissas de **D** não podem ser enfraquecidas ou simplificadas enquanto preservam a validade de **D**. Portanto, para Leplin, o que estabelece se **O** é uma consequência nova de **T** não é saber se **O** foi realmente utilizada na construção de **T**, mas antes se ela foi "requisitada" para a construção de **T**. Assim como a madura concepção de novidade do "uso essencial" de Worrall, o que importa é saber se há um caminho heurístico para **T** que não apele a **O**, tenha ou não **O** sido usada na construção de **T**. A Condição de Singularidade ajuda a reforçar o argumento em favor da verdade de teorias com consequências novas verdadeiras, pois se houvesse outra teoria **T'** (incompatível com **T**) que também fornece uma explicação viável de **O**, a imputação da verdade não poderia explicar o novo sucesso de **T** e de **T***'* simultaneamente. O sucesso de ao menos uma delas teria de ser por acaso, mas se o acaso pudesse explicar um desses sucessos, ele poderia explicar o outro também.

As duas condições para a novidade têm sido questionadas. Dada a Condição de Independência, não é claro que qualquer resultado observacional **O** contará como novo para qualquer teoria, pois pode sempre ser verdadeiro que o conjunto mais fraco de premissas que implicam **T** (que será citado numa reconstrução minimamente adequada do raciocínio que leva a **T**) incluirá **O** como disjunto de uma das premissas (Healey 2001, p. 779). A Condição de Singularidade insiste que não haja explicações alternativas de **O** no momento em que **T** inaugura uma explicação de **O**, embora claramente teorias que explicam **O** pudessem ser propostas e vir não menos a ameaçar a imputação da verdade a **T**. Essas condições parecem arbitrariamente privilegiar teorias dependendo de quando elas foram propostas (SAKAR, 1998, p. 206-208; LADYMAN, 1999, p. 184).

Outra concepção de novidade cuja intenção é reforçar o argumento definitivo em favor do realismo é a "novidade funcional" (ALAI, 2014). Um dado d é 'novo funcionalmente' para uma teoria **T** se (1) **d** não foi utilizado essencialmente na construção de **T** (há um caminho heurístico para **T** e hipóteses auxiliares relacionadas que não citam d), (2) **d** é improvável *a priori*, e (3) **d** é heterogêneo em relação aos dados utilizados para construir **T** e hipóteses auxiliares relacionadas, isto é, **d** é qualitativamente diferente de tais dados. A novidade funcional é um conceito "gradual" na medida em que improbabilidade *a priori* e heterogeneidade dos dados vem em graus. Se houver mais que uma teoria para a qual d é novo funcionalmente, então a disputa entre essas teorias não pode ser resolvida pelo argumento definitivo (ALAI, 2014, p. 306).

Antirrealistas argumentaram que, na medida em que adotamos uma filosofia naturalista da ciência, os mesmos padrões deveriam ser usados para avaliar teorias filosóficas como teorias científicas. Consequentemente, se novas confirmações são necessárias para inferir a verdade da teoria, então o realismo científico não deve ser aceito como verdadeiro, já que a última tese não possui novas confirmações para seu crédito (FROST-ARNOLD, 2010; MIZRAHI, 2012).

Outro componente do debate realista/antirrealista no qual figura de modo importante o apelo para o sucesso novo é o debate acerca da "indução pessimista" (ou "metaindução pessimista"). De acordo com este argumento, a história da ciência é quase inteiramente uma história de teorias julgadas empiricamente bem-sucedidas em seus dias, apenas para subsequentemente se mostrarem inteiramente falsas. Não há razão para pensar que teorias atualmente aceitas sejam algo diferente nesse sentido (LAUDAN, 1981b).

Em resposta, alguns realistas defenderam o "realismo seletivo", o qual concede que, embora a maioria das teorias da história da ciência tenham se provado falsas, algumas delas possuem componentes que foram retidos nas teorias subsequentes - estes tendem a ser os componentes responsáveis pelos novos sucessos. Exemplos putativos desse fenômeno são a teoria térmica do calórico e as teorias ópticas do século XIX (PSILLOS, 1999, cap. 6), ambas as quais foram definitivamente rejeitadas como falsas, mas tiveram componentes retidos em teorias subsequentes; estes (componentes) eram as porções que eram responsáveis por suas novas confirmações[53] . Assim, em linha com o argumento definitivo, a afirmação feita é a de que novos sucessos constituem um argumento sério a favor da verdade do componente da teoria que os geraram. No entanto, antirrealistas responderam citando casos de afirmações teóricas que foram subsequentemente determinadas a serem inteiramente falsas, mas que conseguiram, apesar disso, gerar registros impressionantes de novas predições. Incluem-se aí certas afirmações chave feitas por Johannes Kepler, em seu *Mysterium Cosmographicum* (1596), suposições usadas por Adams e Leverrier na predição da existência e da localização do planeta Netuno (LYONS, 2006) e a astronomia ptolomaica (CARMAN; DÍEZ, 2015).

---

[53] Kitcher (1993, cap. 5) contém um argumento similar, considerando a teoria óptica do século XIX, mas sem enfatizar especificamente o papel do sucesso novo. Para discussão crítica dos problemas relevantes, confira Chang (2003), Stanford (2006), Harker (2010) e Leconte (2017).

Referências

ACHINSTEIN, P. Explanation vs. Prediction: Which Carries More Weight. **Proceedings of the Biennial Meeting of the Philosophy of Science Association**, v. 2, p. 156-164, 1994. DOI: 10.1086/psaprocbienmeetp.1994.2.192926.

ACHINSTEIN, P. **The Book of Evidence**. Oxford: Oxford University Press, 2001. DOI:10.1093/0195143892.001.0001.

AKAIKE, H. Information Theory as an Extension of the Maximum Likelihood Principle. *In:* PETROV, B. N.; CSAKI, F. (ed.). **Second International Symposium on Information Theory.** Budapest: Akademiai Kiado, 1973, p. 267-281.

AKEROYD, F. M. Prediction and the Periodic Table: A Response to Scerri and Worrall. **Journal for General Philosophy of Science**, v. 34, n. 2, p. 337-355, 2003. DOI:10.1023/B:JGPS.0000005277.60641.ca.

ALAI, M. Novel Predictions and the No Miracle Argument. **Erkenntnis**, v. 79, n. 2, p. 297-326, 2014. DOI:10.1007/s10670-013-9495-7.

ALAI, M. The No Miracle Argument and Strong Predictivism vs. Barnes. *In:* MAGNINI, L.; CASADIO, C. (ed.). **Model Based Reasoning in Science and Technology**. Switzerland: Springer International Publishing, 2016. p.541-556. DOI:10.1007/978-3-319-38983-7_30

BAMFORD, G. Popper's Explication of Ad Hocness: Circularity, Empirical Content, and Scientific Practice. **British Journal for the Philosophy of Science**, v. 44, n. 2, p. 335-355, 1993. DOI:10.1093/bjps/44.2.335.

BARNES, E. C. Discussion: Thoughts on Maher's Predictivism. **Philosophy of Science**, v. 63, p. 401-410, 1996a. DOI:10.1086/289918.

BARNES, E. C. Social Predictivism. **Erkenntnis**, v. 45, n. 1, p. 69-89, 1996b. DOI:10.1007/BF00226371.

BARNES, E. C. The Quantitative Problem of Old Evidence. **British Journal for the Philosophy of Science**, v. 50, n. 2, p. 249-264, 1999. DOI:10.1093/bjps/50.2.249.

BARNES, E. C. Neither Truth Nor Empirical Adequacy Explain Novel Success. **Australasian Journal of Philosophy**, v. 80, n. 4, p. 418-431, 2002. DOI:10.1080/713659528.

BARNES, E. C. Predictivism for Pluralists. **British Journal for the Philosophy of Science**, v. 56, n. 3, p. 421-450, 2005a. DOI:10.1093/bjps/axi131.

BARNES, E. C. On Mendeleev's Predictions: Comment on Scerri and Worrall. **Studies in the History and Philosophy of Science**, v. 36, n. 4, p. 801-812, 2005b. DOI:10.1016/j.shpsa.2005.08.005.

BARNES, E. C. **The Paradox of Predictivism.** Cambridge: Cambridge University Press, 2008. DOI:10.1017/CBO9780511487330.

BARNES, E. C. The Roots of Predictivism. **Studies in the History and Philosophy of Science**, v. 45, p. 46-53, 2014. DOI:10.1016/j.shpsa.2013.10.002.

BRUSH, S. G. Prediction and Theory Evaluation: The Case of Light Bending. **Science**, v. 246, n. 4934, p. 1124-1129, 1989. DOI:10.1126/science.246.4934.1124.

BRUSH, S. G. Prediction and Theory Evaluation: Alfvén on Space Plasma Phenomena", **Eos**, v. 71, n. 2, p. 19-33, 1990. DOI:10.1029/EO071i002p00019.

BRUSH, S. G. Prediction and Theory Evaluation: Cosmic Microwaves and the Revival of the Big Bang. **Perspectives on Science**, v. 1, n. 4, p. 565-601,1993.

BRUSH, S. G. Dynamics of Theory Change: The Role of Predictions. **Proceedings of the Biennial Meeting of the Philosophy of Science Association**, v. 2, p. 133-145, 1994. DOI:10.1086/psaprocbienmeetp.1994.2.192924.

BRUSH, S. G. The Reception of Mendeleev's Periodic Law in America and Britain. **Isis**, v. 87, n. 4, p. 595-628,1996. DOI:10.1086/357649.

BRUSH, S. G. Predictivism and the Periodic Table. **Studies in the History and Philosophy of Science Part A**, v. 38, n. 1, p. 256-259, 2007. DOI:10.1016/j.shpsa.2006.12.007.

CAMPBELL, R.; VINCI,T. Novel Confirmation. **British Journal for the Philosophy of Science**, v. 34, n. 4, p. 315-341, 1983. DOI:10.1093/bjps/34.4.315.

CARMAN, C.; DÍEZ, J. Did Ptolemy Make Novel Predictions? Launching Ptolemaic Astronomy into the Scientific Realism Debate. **Studies in the History and Philosophy of Science**, v. 52, p. 20-34, 2015. DOI:10.1016/j.shpsa.2015.04.002.

CARRIER, M. Prediction in context: On the comparative epistemic merit of predictive success. **Studies in the History and Philosophy of Science**, v. 45, p. 97-102, 2014. DOI:10.1016/j.shpsa.2013.10.003.

CHANG, H. Preservative Realism and Its Discontents: Revisiting Caloric. **Philosophy of Science**, v. 70, n. 5, p. 902-912, 2003. DOI:10.1086/377376.

CHRISTIANSEN, D. Measuring Confirmation. **Journal of Philosophy**, v. 96, n.9, p. 437-461, 1999. DOI:10.2307/2564707.

COLLINS, R. Against the Epistemic Value of Prediction over Accommodation. **Noûs**, v. 28, n. 2, p. 210-224, 1994. DOI:10.2307/2216049.

DAWID, R.; HARTMANN, S. The No Miracles Argument without the Base-Rate Fallacy. **Synthese,** 2017. DOI:10.1007/s11229-017-1408-x.

DICKEN, P. Normativity, the Base-Rate Fallacy, and Some Problems for Retail Realism. **Studies in the History and Philosophy of Science Part A**, v. 44, n. 4, p. 563-570, 2013.

DOUGLAS, H.; MAGNUS, P. D. State of the Field: Why Novel Prediction Matters. **Studies in the History and Philosophy of Science**, v. 44, n. 4, p. 580-589, 2013. DOI:10.1016/j.shpsa.2013.04.001.

EELLS, E.; FITELSON, B. Measuring Confirmation and Evidence. **Journal of Philosophy**, v. 97, n. 12, p. 663-672, 2000. DOI:10.2307/2678462.

FORSTER, M. R. Predictive Accuracy as an Achievable Goal of Science. **Philosophy of Science**, v. 69, n. S3, p. S124-S134, 2002. DOI:10.1086/341840.

FORSTER, M.; SOBER, E. How to Tell when Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions. **British Journal for the Philosophy of Science**, v. 45, n. 1, p. 1-35, 1994. DOI:10.1093/bjps/45.1.1.

FRANKEL, H. The Career of Continental Drift Theory: An application of Imre Lakatos' analysis of scientific growth to the rise of drift theory. **Studies in the History and Philosophy of Science**, v. 10, n. 1, p. 21-66, 1979. DOI:10.1016/0039-3681(79)90003-7.

FRISCH, M. Predictivism and Old Evidence: A Critical Look at Climate Model Tuning. **European Journal for the Philosophy of Science**, v. 5, n. 2, p. 171-190, 2015. DOI:10.1007/s13194-015-0110-4.

FROST-ARNOLD, G. The No-Miracles Argument for Scientific Realism: Inference to an Unacceptable Explanation. **Philosophy of Science**, v. 77, n. 1, p. 35-58, 2010. DOI:10.1086/650207.

GARDNER, Michael R. Predicting Novel Facts. **British Journal for the Philosophy of Science**, v. 33, n. 1, p. 1-15, 1982. DOI:10.1093/bjps/33.1.1.

GIERE, R. N. **Understanding Scientific Reasoning**. 2. ed. New York: Holt, Rinehart, e Winston, 1984.

GLYMOUR, C. N. **Theory and Evidence, Princeton.** NJ: Princeton University Press, 1980.

GLYMOUR, C. N. Review: The Paradox of Predictivism by Eric Christian Barnes. **Notre Dame Philosophical Reviews**, 2008. Disponível em: <https://ndpr.nd.edu/news/the-paradox-of-predictivism/>. Acesso em: 29 set. 2021.

GOOD, I. J. The White Shoe is a Red Herring. **British Journal for the Philosophy of Science**, v. 17, n. 4, p. 322, 1967. DOI:10.1093/bjps/17.4.322.

GOODMAN, N. **Fact, Fiction and Forecast.** 4. ed. Cambridge: Harvard University Press, 1983.

GRÜNBAUM, A. Ad Hoc Auxiliary Hypotheses and Falsificationism. **British Journal for the Philosophy of Science**, v. 27, n. 4, p. 329-362, 1976. DOI:10.1093/bjps/27.4.329.

HACKING, I. Imre Lakatos's Philosophy of Science. **British Journal for the Philosophy of Science**, v. 30, n. 4, p. 381-410, 1979. DOI:10.1093/bjps/30.4.381.

HARKER, D. Accommodation and Prediction: The Case of the Persistent Head. **British Journal for the Philosophy of Science**, v. 57, n. 2, p. 309-321, 2006. DOI:10.1093/bjps/axl004.

HARKER, D. The Predilections for Predictions. **British Journal for the Philosophy of Science**, v. 59, n. 3, p. 429-453, 2008. DOI:10.1093/bjps/axn017.

HARKER, D. Two Arguments for Scientific Realism Unified. **Studies in the History and Philosophy of Science**, v. 41, n. 2, p. 192-202, 2010. DOI:10.1016/j.shpsa.2010.03.006.

HARKER, D. Review: The Paradox of Predictivism by Eric Christian Barnes. **British Journal for the Philosophy of Science**, v. 62, n. 1, p. 219-223, 2011. DOI:10.1093/bjps/axq027.

HARTMAN, S.; FITELSON, B. A New Garber-Style Solution to the Problem of Old Evidence. **Philosophy of Science**, v. 82, n. 4, p. 712-717, 2015. DOI:10.1086/682916.

HEALEY, R. Review: A Novel Defense of Scientific Realism by Jarrett Leplin. **Mind**, v. 110, n. 439, p. 777-780, 2001. DOI:10.1093/mind/110.439.777.

HENDERSON, L. The No Miracles Argument and the Base-Rate Fallacy. **Synthese,** v. 4, p. 1295-1302, 2017.

HITCHCOCK, C.; SOBER, E. Prediction versus Accommodation and the Risk of Overfitting. **British Journal for the Philosophy of Science**, v. 55, n. 1, p. 1-34, 2004. DOI:10.1093/bjps/55.1.1.

HOLTON, G. **Thematic Origins of Scientific Thought**: Kepler to Einstein. Ed. rev. Cambridge: Harvard University Press, 1988.

HOWSON, C. Bayesianism and Support by Novel Facts. **British Journal for the Philosophy of Science**, v. 35, n. 3, p. 245-251, 1984. DOI:10.1093/bjps/35.3.245.

HOWSON, C. Accommodation, Prediction and Bayesian Confirmation Theory. **Proceedings of the Biennial Meeting of the Philosophy of Science Association**, v. 2, p. 381-392, 1988. DOI:10.1086/psaprocbienmeetp.1988.2.192899.

HOWSON, C. Fitting Your Theory to the Facts: Probably Not Such a Bad Thing After All. *In:* SAVAGE, C. W. (ed.). **Scientific Theories**. v. 14. Minneapolis: University of Minnesota Press, 1990. p. 224-244. Disponível em: <https://cla.umn.edu/mcps/publications/minnesota-studies-philosophy-science>. Acesso em: 29 set. 2021.

HOWSON, C.; FRANKLIN, A. Maher, Mendeleev and Bayesianism. **Philosophy of Science**, v. 58, n. 4, p. 574-585, 1991. DOI:10.1086/289641.

HUDSON, R. G. Novelty and the 1919 Eclipse Experiments. **Studies in the History and Philosophy of Modern Physics**, v. 34, n. 1, p. 107-129, 2003. DOI:10.1016/S1355-2198(02)00082-5.

HUDSON, R. G. What's Really at Issue with Novel Predictions? **Synthese**, v. 155, n. 1, p. 1-20, 2007. DOI:10.1007/s11229-005-6267-1.

HUNT, J. C. On Ad Hoc Hypotheses. **Philosophy of Science**, v. 79, n. 1, p. 1-14, 2012. DOI:10.1086/663238.

ISEDA, T. Use-Novelty, Severity, and a Systematic Neglect of Relevant Alternatives. **Philosophy of Science**, v. 66, p. S403-S413, 1999. DOI:10.1086/392741.

KAHN, J. A., LANDSBERG, S. E.; STOCKMAN, A. C. On Novel Confirmation. **British Journal for the Philosophy of Science**, v. 43, 1990, p. 503-516.

KEYNES, J. M. **A Treatise on Probability**. London: Macmillan, 1921.

KISH, L. Some Statistical Problems in Research Design. *In:* MORRISON, D. E.; HENKEL, R. E. (ed.). **The Significance Test Controversy**: A Reader. Chicago: Aldine, 1959, p. 127-141. DOI:10.2307/2089381.

KITCHER, P. **The Advancement of Science**: Science without Legend, Objectivity without Illusions. Oxford: Oxford University Press, 1993.

LADYMAN, J. Review: Jarrett Leplin, A Novel Defense of Scientific Realism. **British Journal for the Philosophy of Science**, v. 50, n. 1, p. 181-188, 1999. DOI:10.1093/bjps/50.1.181

LAKATOS, I. Falsification and the Methodology of Scientific Research Programmes. *In:* LAKATOS, I; MUSGRAVE, A. (ed.). **Criticism and the Growth of Knowledge**: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965. Cambridge: Cambridge University Press, 1970, p. 91-196. DOI:10.1017/CBO9781139171434.009.

LAKATOS, I. History of Science and its Rational Reconstructions. *In:* BUCK, R. C.; COHEN, R. S. (ed.). **PSA 1970** (Boston Studies in the Philosophy of Science, 8). Dordrecht: Springer Netherlands, 1971, p. 91-135. DOI:10.1007/978-94-010-3142-4_7.

LANGE, M. The Apparent Superiority of Prediction to Accommodation: a Reply to Maher. **British Journal for the Philosophy of Science**, v. 52, n. 3, p. 575-588, 2001. DOI:10.1093/bjps/52.3.575.

LAUDAN, L. The Epistemology of Light: Some Methodological Issues in the Subtle Fluids Debate. *In:* **Science and Hypothesis: Historical Essays on Scientific Methodology** (University of Western Ontario Series in Philosophy of Science, 19). Dordrecht: D. Reidel, 1981a, p. 111-140.

LAUDAN, L. A Confutation of Convergent Realism. **Philosophy of Science**, v. 48, n. 1, p. 19-49, 1981b. DOI:10.1086/288975.

LECONTE, G. Predictive Success, Partial Truth, and Duhemian Realism. **Synthese**, v. 194, n. 9, p. 3245-3265, 2017. DOI:10.1007/s11229-016-1305-8.

LEE, W. Y. Hitchcock and Sober on Weak Predictivism. **Philosophia**, v. 40, n. 3, p. 553-562, 2012. DOI:10.1007/s11406-011-9331-8.

LEE, W. Y. Akaike's Theorem and Weak Predictivism in Science. **Studies in the History and Philosophy of Science Part A**, v. 44, n. 4, p. 594-599, 2013. DOI:10.1016/j.shpsa.2013.06.001.

LEPLIN, J. The Concept of an Ad Hoc Hypothesis. **Studies in History and Philosophy of Science**, v. 5, n. 3, p. 309-345, 1975. DOI:10.1016/0039-3681(75)90006-0.

LEPLIN, J. The Assessment of Auxiliary Hypotheses**. British Journal for the Philosophy of Science**, v. 33, n. 3, p. 235-249, 1982. DOI:10.1093/bjps/33.3.235.

LEPLIN, J. The Bearing of Discovery on Justification. **Canadian Journal of Philosophy**, v. 17, p. 805-814, 1987. DOI:10.1080/00455091.1987.10715919.

LEPLIN, J. A Novel Defense of Scientific Realism. Oxford: Oxford University Press, 1997.

LEPLIN, J. Review: The Paradox of Predictivism by Eric Christian Barnes. **The Review of Metaphysics**, v. 63, n. 2, p. 455-457, 2009.

LIPTON, P. Prediction and Prejudice. **International Studies in the Philosophy of Science**, v. 4, n. 1, p. 51-65, 1990. DOI:10.1080/02698599008573345.

LIPTON, P. **Inference to the Best Explanation**. London/New York: Routledge, 1991.

LYONS, T. D. Scientific Realism and the Strategema de Divide et Impera. **British Journal for the Philosophy of Science**, v. 57, n. 3, p. 537-560, 2006. DOI:10.1093/bjps/axl021.

MAGNUS, P.D. Miracles, trust, and ennui in Barnes' Predictivism. **Logos & Episteme**, v. 2, n. 1, p. 103-115, 2011. DOI:10.5840/logos-episteme20112152.

MAGNUS, P. D.; CALLENDER, C. Realist Ennui and the Base Rate Fallacy. **Philosophy of Science**, v. 71, n. 3, p. 320-338, 2004. DOI:10.1086/421536.

MAHER, P. Prediction, Accommodation, and the Logic of Discovery. **Proceedings of the Biennial Meeting of the Philosophy of Science Association,** v. 1, p. 273-285, 1988. DOI:10.1086/psaprocbienmeetp.1988.1.192994.

MAHER, P. How Prediction Enhances Confirmation. *In:* DUNN, J. M.; GUPTA, A. (ed.). **Truth or Consequences:** Essays in Honor of Nuel Belnap. Dordrecht: Kluwer, 1990, p. 327-343.

MAHER, P. Howson and Franklin on Prediction. **Philosophy of Science**, v. 60, n. 2, p. 329-340, 1993. DOI:10.1086/289736.

MAYO, D. G. Novel Evidence and Severe Tests. **Philosophy of Science**, v. 58, n. 4, p. 523-552, 1991. DOI:10.1086/289639.

MAYO, D. G. **Error and the Growth of Experimental Knowledge**. London: University of Chicago Press, 1996.

MAYO, D. G. Novel Work on the Problem of Novelty? Comments on Hudson. **Studies in the History and Philosophy of Modern Physics**, v. 34, p. 131-134, 2003. DOI:10.1016/S1355-2198(02)00083-7.

MAYO, D. G. How to Discount Double-Counting When It Counts: Some Clarifications. **British Journal for the Philosophy of Science**, v. 59, n. 4, p. 857-879, 2008. DOI:10.1093/bjps/axn034.

MAYO, D. G. An Ad Hoc Save of a Theory of Adhocness? Exchanges with John Worrall. **Mayo e Spanos**, 2010, p. 155-169.

MAYO, D. G. Some surprising facts about (the problem of) surprising facts (from the Dusseldorf Conference, February 2011). **Studies in the History and Philosophy of Science**, v. 45, p. 79-86, 2014. DOI:10.1016/j.shpsa.2013.10.005.

MAYO, D. G.; SPANOS, A. (ed.). **Error and Inference**: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science. Cambridge: Cambridge University Press, 2010. DOI:10.1017/CBO9780511657528.

MCCAIN, K. A Predictivist Argument Against Skepticism. **Analysis**, v. 72, n. 4, p. 660-665, 2012. DOI:10.1093/analys/ans109.

MCINTYRE, L. Accommodation, Prediction, and Confirmation. **Perspectives on Science**, v. 9, n. 3, p. 308-328, 2001. DOI:10.1162/10636140160176161.

MENKE, C. Does the Miracle Argument Embody a Base-Rate Fallacy?. **Studies in the History and Philosophy of Science Part A**, v. 45, p. 103-108, 2014.

MILL, J. S. **A System of Logic, Ratiocinative and Inductive**: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. v. 2. London: John W. Parker, 1843.

MIZRAHI, M. Why the Ultimate Argument for Scientific Realism Fails. **Studies in the History and Philosophy of Science**, v. 43, n. 1, p. 132-138, 2012. DOI:10.1016/j.shpsa.2011.11.001.

MURPHY, N. Another Look at Novel Facts. **Studies in the History and Philosophy of Science**, v. 20, n. 3, p. 385-388, 1989. DOI:10.1016/0039-3681(89)90014-9.

MUSGRAVE, A. Logical versus Historical Theories of Confirmation. **British Journal for the Philosophy of Science**, v. 25, n. 1, p. 1-23, 1974. DOI:10.1093/bjps/25.1.1.

MUSGRAVE, A. The Ultimate Argument for Scientific Realism. *In:* NOLA, R. (ed.). **Relativism and Realism in Science.** Dordrecht: Kluwer Academic Publishers, 1988, p. 229-252. DOI:10.1007/978-94-009-2877-0_10.

NUNAN, R. Novel Facts, Bayesian Rationality, and the History of Continental Drift. **Studies in the History and Philosophy of Science**, v. 15, n. 4, p. 267-307, 1984. DOI:10.1016/0039-3681(84)90013-X.

PARTINGTON, J. R.; MCKIE, D. Historical Studies on the Phlogiston Theory: I. The Levity of Phlogiston. **Annals of Science**, v. 2, n. 4, p. 361-404, 1936. DOI:10.1080/00033793700200691.

PARTINGTON, J. R.; MCKIE, D. Historical Studies on the Phlogiston Theory: II. The Negative Weight of Phlogiston, **Annals of Science,** v. 3, n.1, p. 1-58, 1938a. DOI:10.1080/00033793800200781.

PARTINGTON, J. R.; MCKIE, D. Historical Studies on the Phlogiston Theory: III. Light and Heat in Combustion, **Annals of Science,** v. 3, n. 4, p. 337-371, 1938b. DOI:10.1080/00033793800200951.

POPPER, K. **Conjectures and Refutations**: The Growth of Scientific Knowledge, New York: Harper and Row, 1963.

POPPER, K. **Objective Knowledge.** Oxford: Clarendon Press, 1972.

POPPER, K. Replies to my critics. *In:* SCHILPP, P. A. (ed.). **The Philosophy of Karl POPPER:** Book II, 961–1197 La Salle**.** Illinois: Open Court, 1974.

PSILLOS, S. **Scientific Realism**: How Science Tracks the Truth, London and New York: Routledge, 1999.

PUTNAM, H. **Philosophical Papers.** Cambridge: Cambridge University Press, 1975. v. 1.

REDHEAD, M. Adhocness and the Appraisal of Theories. **British Journal for the Philosophy of Science**, v. 29, p. 355-361, 1978.

SALMON, W. C. Rational Prediction. **British Journal for the Philosophy of Science**, v. 32, n. 2, p. 115-125, 1981. DOI:10.1093/bjps/32.2.115.

SARKAR, H. Review of A Novel Defense of Scientific Realism by Jarrett Leplin. **Journal of Philosophy**, v. 95 n. 4, p. 204-209, 1998. DOI:10.2307/2564685.

SCERRI, E. R. Response to Barnes's critique of Scerri and Worrall. **Studies in the History and Philosophy of Science**, v. 36, n. 4, p. 813-816, 2005. DOI:10.1016/j.shpsa.2005.08.006.

SCERRI, E. R.; WORRALL, J. Prediction and the Periodic Table. **Studies in the History and Philosophy of Science**, v. 32,n. 3, p. 407-452, 2001. DOI:10.1016/S0039-3681(01)00023-1.

SCHINDLER, S. Use Novel Predictions and Mendeleev's Periodic Table: Response to Scerri and Worrall (2001). **Studies in the History and Philosophy of Science Part A**, v. 39, n. 2, p. 265-269, 2008. DOI:10.1016/j.shpsa.2008.03.008.

SCHINDLER, S. Novelty, coherence, and Mendeleev's periodic table. **Studies in the History and Philosophy of Science Part A**, v. 45, p. 62-69, 2014. DOI:10.1016/j.shpsa.2013.10.007.

SCHLESINGER, G. N. Accommodation and Prediction. **Australasian Journal of Philosophy**, v. 65, n. 1, p. 33–42, 1987. DOI:10.1080/00048408712342751.

SCHURZ, G. Bayesian Pseudo-Confirmation, Use-Novelty, and Genuine Confirmation. **Studies in History and Philosophy of Science Part A**, v. 45, p. 87-96, 2014. DOI:10.1016/j.shpsa.2013.10.008.

STANFORD, P. K. **Exceeding Our Grasp**: Science, History, and the Problem of Unconceived Altenatives. Oxford: Oxford University Press, 2006. DOI:10.1093/0195174089.001.0001.

SWINBURNE, R. **Epistemic Justification.** Oxford: Oxford University Press, 2001. DOI:10.1093/0199243794.001.0001.

THOMASON, N. Could Lakatos, Even with Zahar's Criterion of Novel Fact, Evaluate the Copernican Research Programme? **British Journal for the Philosophy of Science**, v. 43, n. 2, p. 161-200, 1992. DOI:10.1093/bjps/43.2.161.

VOTSIS, I. Objectivity in Confirmation: Post Hoc Monsters and Novel Predictions. **Studies in the History and Philosophy of Science Part A**, v. 45, p. 70-78, 2014. DOI:10.1016/j.shpsa.2013.10.009.

WHEWELL, W. **Mr. Mill's Logic**. Pittsburgh: University of Pittsburgh Press, 1968, p. 265-308.

WHITE, R. The Epistemic Advantage of Prediction over Accommodation. **Mind**, v. 112, n. 448, p. 653-683, 2003. DOI:10.1093/mind/112.448.653.

WORRALL, J. The Ways in Which the Methodology of Scientific Research Programmes Improves Upon Popper's Methodology. *In:* GERARD R.; GUNNAR, A. (ed.). **Progress and Rationality in Science**. Dordrecht: D. Reidel, 1978, p. 45-70, 1978. DOI:10.1007/978-94-009-9866-7_3.

WORRALL, J. Scientific Discovery and Theory-Confirmation. *In:* PITT, J. C. (ed.). **Change and Progress in Modern Science**: Papers Related to and Arising from the Fourth International Conference on History and Philosophy of Science, Blacksburg, Virginia, November 1982. Dordrecht: D. Reidel, 1985, p. 301-331. DOI:10.1007/978-94-009-6525-6_11.

WORRALL, J. Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories. *In:* GOODING, D.; PINCH, T.; SCHAFFER, S. (eds.). **The Uses of Experiment**: Studies in the Natural Sciences. Cambridge: Cambridge University Press, 1989, p. 135-157.

WORRALL, J. New Evidence for Old. *In:* GÄRDENFORS, P.; WOLENSKI, J.; KIJANIA-PLACEK, K. (ed.) **In the Scope of Logic, Methodology and Philosophy of Science**: Volume One of the 11th International Congress of Logic, Methodology and Philosophy of Science, Cracow, August 1999. Dordrecht: Kluwer Academic Publishers, 2002, p. 191-209.

WORRALL, J. Prediction and the 'Periodic Law': A Rejoinder to Barnes. **Studies in the History and Philosophy of Science**, v. 36, n. 4, p. 817-826, 2005. DOI:10.1016/j.shpsa.2005.08.007.

WORRALL, J. Theory-Confirmation and History. *In:* CHEYNE, C.; WORRALL. J. (ed.). **Rationality and Reality**: Conversations with Alan Musgrave. Dordrecht: Springer, 2006, p. 31-61. DOI:10.1007/1-4020-4207-8_4.

WORRALL, J. Error, tests and theory confirmation. *In:* MAYO, D. G.; SPANOS, Aris, (ed.) **Error and Inference**: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science. Cambridge: Cambridge University Press, 2010, p. 125-154.

WORRALL, J. Prediction and Accommodation Revisited. **Studies in History and Philosophy of Science Part A**, v. 45, p. 54-61, 2014. DOI: 10.1016/j.shpsa.2013.10.001.

WRIGHT, J. **Explaining Science's Success**: Understanding How Scientific Knowledge Works, Durham. England: Acumen, 2012.

ZAHAR, E. Why did Einstein's Programme supercede Lorentz's? (I). **British Journal for the Philosophy of Science**, v. 24, n. 2, p. 95-123, 1973. DOI:10.1093/bjps/24.2.95.

ZAHAR, E. **Einstein's Revolution**: A Study In Heuristic. La Salle: Open Court, 1983.

(V) Representação Científica*

Autores: Roman Frigg e James Nguyen
Tradução: Ana Margarete Barbosa de Freitas
Revisão: Nathália C A. P. Strogen

A ciência nos fornece representações de átomos, partículas elementares, polímeros, populações, árvores genéticas, economias, decisões racionais, aviões, terremotos, incêndios florestais, sistemas de irrigação e sobre o clima do mundo. É através destas representações que aprendemos sobre o mundo. Este verbete explora várias explicações diferentes de representação científica, com um foco particular em como os **modelos** científicos representam seus sistemas-alvo. Como os filósofos da ciência estão gradativamente reconhecendo a importância, se não a primazia, dos modelos científicos como unidades representacionais da ciência, é importante enfatizar que a forma como eles representam desempenha um papel

---

* FRIGG, R; NGUYEN, J. Scientific Representation. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2020. Disponível em: https://plato.stanford.edu/archives/spr2020/entries/scientific-representation/. Acesso em: 22 set. 2021.

The following is the translation of the entry on Scientific Representation by Roman Frigg e James Nguyen, 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/spr2020/entries/scientific-representation/. 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/scientific-representation/. 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.

fundamental na maneira como devemos responder a outras perguntas na filosofia da ciência (como por exemplo, o debate sobre o realismo científico). Este verbete começa por desvendar "o" problema da representação científica e, em seguida, avalia criticamente as opções atuais disponíveis na literatura.

1. Problemas relacionados à representação científica

Em termos mais gerais, qualquer representação que é o produto de um empreendimento científico é uma representação científica. Essas representações são um grupo heterogêneo que compreende desde leituras de termômetro e fluxogramas até descrições verbais, fotografias, imagens de raios-x, imagens digitais, equações, modelos e teorias. Como essas representações funcionam?

A primeira coisa que impressiona o iniciante no debate sobre representação científica é que parece haver pouco consenso sobre qual é o problema. Diferentes autores enquadram o problema da representação científica de diferentes maneiras e, eventualmente, examinam diferentes questões. Portanto, uma discussão sobre representação científica deve começar com um esclarecimento do problema em si. A revisão de literatura sobre o assunto nos leva à conclusão de que não há tal coisa como o problema da representação científica – de fato, existem pelo menos cinco problemas diferentes relacionados à representação científica. Nesta seção, formularemos estes problemas e articularemos cinco condições de adequação que toda explicação sobre representação científica deve satisfazer.

O primeiro problema é: o que transforma algo em uma representação científica de alguma outra coisa? Tornou-se costumeiro formular esse problema em termos de condições necessárias e suficientes e perguntar: o que preenche o espaço em branco em "**S** é uma representação científica de **T** sse ___", onde "**S**" significa o objeto que representa e "**T**" o "sistema-alvo", a parte ou aspecto do mundo a qual a representação se refere?[54]

---

[54] Diferentes formulações desse problema podem ser encontradas em Frigg (2002, p. 2- 17), Morrison (2008, p. 70) e Suárez (2003, p. 230), e muitos contribuintes para o debate pressupõem tacitamente uma formulação do problema em termos de condições necessárias e suficientes. A análise de conceitos em termos de condições necessárias e suficientes não tem tido uma boa impressão em outros aspectos da filosofia; ver Laurence e Margolis (1999) para uma revisão e discussão. Limitações de espaço nos impedem de prosseguir com esse problema. Queremos sinalizar, no entanto, que se os argumentos padrões contra tal análise ficarem restrito a casos centrais, eles podem perder muito de sua força.

Vamos chamá-lo de *Problema da Representação Científica* (ou Problema RC, abreviado).

Vários colaboradores do debate têm enfatizado que a representação científica é um conceito intencional, que depende de fatores como intenções, propósitos e objetivos de um usuário, padrões contextuais de precisão e públicos-alvo (GIERE, 2010; MÄKI, 2011; SUÁREZ, 2004; VAN FRAASSEN, 2008). Discutiremos isso em detalhes abaixo. Neste ponto, é preciso enfatizar que o enquadramento do problema em termos de uma bicondicional não impede que esses fatores façam parte da análise. Talvez surja a interpretação de que a formulação acima represente uma relação intrínseca entre **S** e **T**, isto é, uma relação que depende apenas das propriedades intrínsecas de **S** e **T**. Essa é uma interpretação errônea. O espaço em branco pode ser preenchido com uma relação **C** *(***S***,* **T***,* $\mathbf{x_1}$*,…,* $\mathbf{x_n}$*)* (**C** no sentido de "constitui") binária (n + 2), onde n ≥ 0 é um número natural e $\mathbf{x_i}$ são fatores como intenções e propósitos.

Uma primeira condição importante de adequação a qualquer resposta a esse problema é que as representações científicas nos permitem formar hipóteses sobre seus sistemas-alvo. Uma imagem de raio-x fornece informações sobre os ossos do paciente, e os modelos permitem que os investigadores descubram características das coisas que os modelos representam. Toda teoria aceitável da representação científica deve explicar como o raciocínio conduzido nas representações pode gerar afirmações sobre seus sistemas-alvo. Swoyer (1991, p. 449) refere-se a esse tipo de pensamento baseado em representação como "raciocínio substitutivo" e, portanto, chamaremos isso de *Condição de Raciocínio Substitutivo*[55] . Essa condição distingue modelos de representações lexicográficos e indexais, que não permitem o raciocínio substitutivo.

Infelizmente, essa condição não restringe as respostas o suficiente, porque qualquer descrição da representação que preencha o espaço em branco de uma maneira que satisfaça a condição de raciocínio substitutivo, invariavelmente, também

---

[55] Parece haver um amplo consenso sobre esse ponto. Bolinska (2013) e Contessa (2007) referem-se a modelos de “representações epistêmicas” e Bailer-Jones (2003, p. 59), Frigg (2006, p. 51), Liu (2013, p. 93), Morgan e Morrison (1999, p. 11); Suárez (2003, p. 229) e Weisberg (2013, p. 150) enfatizaram, de maneiras diferentes, que estudar uma representação científica deve ser capaz de nos fornecer informações sobre a natureza de um sistema-alvo. Uma exceção digna de nota é Callender e Cohen (2006); discutiremos seus pontos de vista na Seção 2.

abrangerá outros tipos de representações. As fotografias dos radares de trânsito fornecem à polícia informações sobre motoristas que infringem a lei, um modelo de papelão de um castelo nos instrui sobre seu *layout* e proporções, e um mapa do tempo nos mostra onde devemos esperar que chova e, portanto, é provável que caiam sob uma descrição de representação que explica o raciocínio substitutivo. Portanto, outras representações que não sejam científicas também permitem raciocínio substitutivo, o que levanta a questão: como as representações científicas diferem de outros tipos de representações que permitem raciocínio substitutivo? Callender e Cohen (2006, p. 68-69) apontam que esta é uma versão do problema da demarcação de Popper, agora redigido em termos de representação e, por isso, nos referimos a ele como o *Problema da Demarcação Representacional*.

Callender e Cohen expressam ceticismo quanto à existência de uma solução para esse problema e sugerem que a distinção entre representações científicas e não científicas é circunstancial (2006: 83): representações científicas são representações usadas ou desenvolvidas por alguém que é um cientista. Outros autores não discutem explicitamente o problema da demarcação representacional, mas posturas semelhantes às de Callender e Cohen estão implícitas em qualquer abordagem que analise a representação científica juntamente com outros tipos de representação. Elgin (2010), French (2003), Frigg (2006), Hughes (1997), Suárez (2004) e van Fraassen (2008), por exemplo, todos traçam paralelos entre representação científica e pictórica, o que faria pouco sentido se as representações pictórica e científica fossem categoricamente diferentes.

Quem rejeita a noção de que existe uma diferença essencial entre representação científica e não científica pode seguir uma sugestão de Contessa (2007) e ampliar o escopo da investigação. Ao invés de analisar a representação científica, eles podem analisar a categoria mais ampla de **representação epistêmica**. Essa categoria compreende representações científicas, mas também inclui outras representações que permitem o raciocínio substitutivo. A tarefa passa a ser preencher o espaço em branco em "**S** é uma representação epistêmica de **T** sse ___". Chamamos isso de *Problema da Representação Epistêmica* (Problema RE, para abreviar), e a bicondicional de Esquema RE. Então, pode-se dizer que o Problema RE é preencher o espaço em branco no Esquema RE.

Nem todas as representações são do mesmo tipo, nem mesmo se restringirmos nossa atenção às representações científicas (assumindo que elas sejam relevantemente diferentes das representações epistêmicas não científicas). Uma fotografia de raio-x representa uma articulação do tornozelo de uma maneira

diferente de um modelo biomecânico, um termômetro de mercúrio representa a temperatura do gás de uma maneira diferente da mecânica estatística, e a teoria química representa um fulereno C60 de maneira diferente da imagem de uma molécula em um microscópio eletrônico. Mesmo ao restringir a atenção ao mesmo tipo de representação, existem diferenças importantes: o modelo de gota líquida de Weizsäcker, por exemplo, representa o núcleo de um átomo de uma maneira que parece ser diferente do modelo de concha, e o modelo de um circuito elétrico representa a função cerebral de uma maneira diferente da de um modelo de rede neural. Em resumo, parece haver diferentes estilos representacionais. Isso levanta a questão: quais estilos existem e como eles podem ser caracterizados?[56] Chamamos isso de *Problema do Estilo*.[57] Não há nenhuma expectativa de que uma lista completa de estilos seja fornecida. De fato, é improvável que essa lista possa ser elaborada e novos estilos serão inventados à medida que a ciência progride. Por esse motivo, uma resposta ao problema do estilo estará sempre aberta, fornecendo uma taxonomia do que está disponível no momento e deixando espaço para adições posteriores.

Algumas representações são precisas; outras não. O modelo da mecânica quântica é uma representação precisa do átomo (ao menos pelo que sabemos até agora), mas o modelo Thomson não é. Com que base fazemos esses julgamentos? Morrison (2008, p. 70) nos lembra que é uma tarefa da teoria da representação identificar o que constitui uma representação precisa. Chamamos isso de problema dos *Padrões de Precisão*. Fornecer esses padrões é um dos problemas que uma explicação sobre a representação precisa resolver.

Esse problema anda de mãos dadas com uma condição de adequação: a *Possibilidade de Deturpação*. Perguntar o que faz de uma representação uma

---

[56] Os estilos podem ser explicados no Esquema RE de pelo menos três maneiras: (i) tendo um esquema diferente para cada tipo de representação; (ii) permitindo que C seja uma relação disjuntiva em que cada disjunção descreva um estilo particular; ou (iii) encontrando um C determinável que tem os diferentes estilos como seus determinantes. Para uma discussão sobre essas opções, consulte (FRIGG; NGUYEN, no prelo, seção 2).

[57] Ver Frigg (2006, p. 50). Mesmo que o problema do estilo não tenha recebido muita atenção explícita, a literatura sobre modelos está repleta de noções que podem ser interpretadas como contribuições para o problema do estilo – modelos analógicos, modelos idealizados e modelos caricaturais são apenas alguns exemplos óbvios. Para uma pesquisa, consulte o verbete Modelos na Ciência.

representação precisa *ipso facto* pressupõe que representações imprecisas também sejam representações. E é assim que deve ser. Se *S* não representar *T* com precisão, será uma deturpação, mas não uma não representação. Então, é uma restrição geral a uma teoria da representação científica que ela tenha de oferecer a possibilidade de deturpação (FRIGG, 2002, p. 16-17; SUÁREZ, 2003, p. 233-235; CONTESSA, 2007; VAN FRAASSEN, 2008, p. 13-15).

Uma condição relacionada diz respeito a modelos que deturpam no sentido de que eles não possuem sistemas-alvo. Modelos de éter, flogisto, populações de quatro sexos etc. são todos considerados modelos científicos, mas não existem éter, flogisto e populações de quatro sexos. Tais modelos carecem de sistemas-alvo (reais), e espera-se que uma explicação da representação científica nos permita entender como esses modelos funcionam. Isso não precisa implicar a afirmação de que são representações no mesmo sentido que modelos com alvos reais e, como discutiremos abaixo, existem explicações que negam aos modelos sem alvo o *status* de representações.

Uma condição adicional de adequação para uma explicação da representação científica é que ela deve explicar a direcionalidade da representação. Como Goodman aponta (1976, p. 5), as representações são sobre seus alvos, mas (pelo menos em geral) os alvos não são sobre suas representações: uma fotografia representa as rachaduras na asa do avião, mas a asa não representa a fotografia. Portanto, há uma direcionalidade essencial nas representações, e uma explicação da representação científica ou epistêmica deve identificar a raiz dessa direcionalidade. Chamamos isso de *Requisito de Direcionalidade*.

Algumas representações, em particular modelos e teorias, são matematizadas e seus aspectos matemáticos são cruciais para sua função cognitiva e representacional. Isso nos força a reconsiderar um quebra-cabeça filosófico consagrado pelo tempo: a aplicabilidade da matemática nas ciências empíricas. O problema pode ser rastreado, pelo menos, até o *Timeu* de Platão, mas sua expressão moderna se deve a Wigner, que nos desafiou a encontrar uma explicação para a enorme utilidade da matemática nas ciências (1960, p. 2). A questão de como um modelo matematizado representa seu alvo implica a questão de como a matemática se aplica a um sistema físico. Consulte Pincock (2012) para uma discussão explícita da relação entre representação científica e a aplicabilidade da matemática. Por esse motivo, nossa quinta e última condição de adequação é que uma explicação da representação precisa explicar como a matemática é aplicada ao mundo físico. Chamamos isso de *Aplicabilidade da Condição Matemática*.

Ao responder às perguntas acima, invariavelmente, nos deparamos com um problema adicional, o *Problema da Ontologia*: que tipos de objetos são representações? Se as representações são objetos materiais, a resposta é direta: placas fotográficas, pedaços de papel cobertos de tinta, blocos elípticos de madeira imersos em água e assim por diante. Mas nem todas as representações são assim. Como Hacking (1983, p. 216) coloca, algumas representações são mantidas mais na cabeça do que nas mãos. O modelo newtoniano do sistema solar, o modelo de Lotka-Volterra de interação predador-presa e a teoria geral da relatividade não são coisas que você pode colocar na mesa do laboratório e observar. O problema da ontologia é esclarecer nossos compromissos e fornecer uma lista com coisas que reconhecemos, ou não, como entidades que desempenham uma função representacional e dão uma explicação do que elas são, no caso dessas entidades levantarem questões (o que exatamente queremos dizer com algo que alguém segura com a cabeça e não com as mãos?). Contessa (2010), Frigg (2010a, b), Godfrey-Smith (2006), Levy (2015), Thomson-Jones (2010), Weisberg (2013), entre outros, chamaram a atenção para esse problema de diferentes maneiras.

Em suma, uma teoria da representação científica deve responder aos seguintes problemas:

1. Abordar o *Problema da Demarcação Representacional* (a questão de como as representações científicas diferem de outros tipos de representações).
2. Aqueles que diferenciam as representações científicas de não científicas devem fornecer uma resposta ao *Problema da Representação Científica* (preenchendo o espaço em branco em "**S** é uma representação científica de **T** sse ___"). Aqueles que rejeitam o problema da demarcação representacional podem resolver o *Problema da Representação Epistêmica* (preenchendo o espaço em branco no Esquema RE: "**S** é uma representação epistêmica de **T** sse ___").
3. Responder ao *Problema do Estilo* (quais estilos existem e como podem ser caracterizados?).
4. Formular *Padrões de Precisão* (como identificamos o que constitui uma representação precisa?).

> 5. Abordar o *Problema da Ontologia* (quais são os tipos de objetos que servem como representações?).

Qualquer resposta satisfatória para essas cinco questões deverá atender às cinco condições de adequação a seguir:

> a. *Raciocínio Substitutivo* (representações científicas nos permitem gerar hipóteses sobre seus sistemas-alvo).
> b. *Possibilidade de Deturpação* (se **S** não representar com precisão **T**, será uma deturpação, mas não uma não representação).
> c. *Modelos sem Alvo* (o que devemos fazer das representações científicas que não têm alvos?).
> d. *Requisito de Direcionalidade* (representações científicas são sobre seus alvos, mas alvos não são sobre suas representações).
> e. *Aplicabilidade Matemática* (como o aparato matemático usado em algumas representações científicas liga-se ao mundo físico).

Listar os problemas dessa maneira não significa dizer que são problemas separados e não relacionados. Essa divisão é analítica, não factual. Serve para estruturar a discussão e avaliar propostas; isso não implica que uma resposta a uma dessas perguntas possa ser dissociada de qual posição adotamos em relação a outras questões.

Qualquer tentativa de resolver essas questões enfrenta um problema metodológico imediato. De acordo com o problema do estilo, existem diferentes tipos de representações: modelos científicos, teorias, resultados de medições, imagens, gráficos, diagramas e asserções linguísticas são representações científicas e, mesmo dentro desses grupos, pode haver uma variação considerável. Mas toda análise tem que começar em algum lugar, o problema é onde. Pode-se adotar uma posição universalista, sustentando que a diversidade de estilos se dissolve durante a análise e, no fundo, todas as instâncias de representação científica/epistêmica funcionam da mesma maneira e são cobertas pela mesma explicação abrangente. Para tal universalista, o problema desaparece porque qualquer ponto de partida leva ao mesmo resultado. Aqueles de tendência particularista negam que exista tal

teoria. Primeiro, eles dividem as representações científicas/epistêmicas em subclasses relevantes e depois analisam cada subclasse separadamente.

Autores diferentes assumem posições diferentes neste debate, e discutiremos suas posições abaixo. Contudo, existem poucos universalistas extremos, se é que existem, e uma revisão como a atual tem que discutir casos diferentes. Infelizmente, as restrições de espaço nos impedem de examinar todas as diferentes variedades de representação científica/epistêmica, e uma seleção deve ser feita. Isso invariavelmente leva à negligência de alguns tipos de representações, e o melhor que podemos fazer sobre isso é ser explícito sobre nossas escolhas. Resolvemos nos concentrar em modelos científicos e, portanto, substituímos nossa variável **S** pelo objeto que representa a variável **M** no modelo. Isso está de acordo com a literatura mais recente sobre representação científica, que se preocupa, predominantemente, com modelos científicos e com a importância primordial que a filosofia da ciência atual atribui aos modelos (*vide* **Modelos em Ciência**).[58]

No entanto, vale a pena mencionar brevemente algumas das omissões que isso traz consigo. Vários tipos de imagens têm seu lugar na ciência, assim como gráficos, diagramas e desenhos. Perini (2010) e Elkins (1999) fornecem discussões sobre representação visual na ciência. As medições também fornecem representações de processos na natureza, às vezes juntamente com a subsequente condensação dos resultados da medição na forma de gráficos, curvas, tabelas e similares. Além disso, as teorias representam o tema com o qual trabalham. Neste ponto, o problemático problema da natureza das teorias volta a surgir. Para uma extensa discussão, confira Portides (no prelo). Os proponentes da visão semântica das teorias interpretam as teorias como famílias de modelos e, portanto, para eles a questão de como as teorias representam coincide com a questão de como os modelos representam. Por outro lado, aqueles que consideram as teorias como entidades linguísticas veem a representação teórica como um tipo especial de

---

[58] Nossa discussão não tem como premissa a afirmação de que todos os modelos são representacionais; nem assume que a representação seja a única função (ou mesmo primária) dos modelos. Tem sido enfatizado de várias maneiras no debate que os modelos desempenham várias funções além da representação; ver Knuuttila (2005, 2011), Morgan e Morrison (1999), Hartmann (1995) e Peschard (2011) para a discussão de diferentes usos não representacionais de modelos. Nossa questão é como os modelos representam quando eles representam.

representação linguística e se concentram na análise de linguagens científicas, em particular a semântica dos chamados termos teóricos.

Antes de aprofundar a discussão, um equívoco comum precisa ser dissipado. O equívoco é que uma representação é uma imagem espelhada, uma cópia ou uma imitação da coisa que representa. Nesta visão, a representação é uma representação realista *ipso facto*. Isto é um erro. As representações podem ser realistas, mas não precisam. E representações certamente não precisam ser cópias da coisa real (uma observação explorada por Lewis Carroll e Jorge Luis Borges em suas sátiras, *Sylvie and Bruno* e *On Exactitude in Science*, respectivamente, sobre cartógrafos que produzem mapas tão grandes quanto o próprio país apenas para vê-los abandonados). Ao longo desta revisão, encontramos posições que abrem espaço para a representação não realista e, portanto, testemunham o fato de que a representação é uma noção muito mais ampla do que o espelhamento.[59]

2. Griceanismo Geral e Decreto Estipulativo

Callender e Cohen (2006) dão uma resposta radical ao problema da demarcação: não há diferença entre representações científicas e outros tipos de representações, nem mesmo entre representações científicas e artísticas. Subjacente a essa reivindicação, há uma posição que eles chamam de "Griceanismo Geral" (GG). O núcleo do GG é a afirmação redutiva de que todas as representações devem seu status como representações a um núcleo privilegiado de representações fundamentais. O GG vem com uma receita prática sobre como proceder com a análise de uma representação:

---

[59] Há, entretanto, outro sentido, no qual pressupomos uma forma mínima de realismo. Ao longo de nossa discussão, assumimos que os sistemas-alvo existem independentemente dos observadores humanos e que são como são, independentemente do que as pessoas pensem sobre eles. Esta é uma pressuposição que os construtivistas (e outros antirrealistas metafísicos) negam. Lynch e Wooglar (1990), por exemplo, argumentam que as representações constituem os fenômenos que representam. Uma avaliação do programa construtivista está além do escopo desta revisão. Vale a pena observar, porém, que pelo menos algumas das questões que discutiremos a seguir reaparecem em uma estrutura construtivista sob o disfarce do problema de como diferentes representações se relacionamento entre si.

> a visão do Griceanismo Geral consiste em dois estágios. Primeiro, explica os poderes representacionais das representações derivadas em termos dos poderes das representações fundamentais; segundo, oferece outra história para explicar a representação para os portadores fundamentais do conteúdo (2006, p. 73).

Destas etapas, apenas a segunda exige um trabalho filosófico sério, e esse trabalho é realizado na filosofia da mente, porque a forma fundamental de representação é a representação mental.

A representação científica é um tipo derivado de representação (2006, p. 71, 75) e, portanto, se enquadra no primeiro estágio da fórmula acima. É reduzido à representação mental por um ato de estipulação. No exemplo de Callender e Cohen, o saleiro na mesa de jantar pode representar Madagascar, desde que alguém estipule que o primeiro represente o segundo, uma vez que

> os poderes representacionais dos estados mentais são tão amplos que podem gerar outras relações representacionais entre relatos arbitrários por força da mera estipulação (2006, pp. 73-74).

Portanto, explicar qualquer forma de representação que não seja a representação mental é uma trivialidade – basta um ato de "decreto estipulativo" (2006: 75). Isso fornece uma resposta para o problema RE:

> *Decreto Estipulativo:* Um modelo científico **M** representa um sistema-alvo **T** sse um usuário do modelo estipula que **M** representa **T**.

O primeiro problema enfrentado pelo *Decreto Estipulativo* é se a estipulação ou as intenções simples dos usuários da linguagem são suficientes para estabelecer relacionamentos representacionais. Na filosofia da linguagem, isso é chamado de problema "Humpty Dumpty". Trata-se de saber se *Humpty Dumpty*, de Lewis Carroll, poderia ou não usar a palavra "glória" para significar "um bom argumento de nocaute" (DONNELLAN, 1968; MACKAY, 1968) (aqui ignoramos a diferença entre significado e denotação). Nesse contexto, não parece que ele pode, e questões

análogas podem ser colocadas no contexto da representação científica: um cientista pode fazer qualquer modelo representar qualquer alvo simplesmente estipulando que isso é assim?

Mesmo que a estipulação fosse suficiente para estabelecer algum tipo de relacionamento representacional, o *Decreto Estipulativo* falha em atender à Condição de Raciocínio Substitutivo: supondo que um saleiro represente Madagascar em virtude do fato de que alguém o estipulou assim, isso não nos diz nada sobre como o saleiro poderia ser usado para aprender sobre Madagascar da maneira que modelos científicos são usados para aprender sobre seus alvos (LIU, 2015, p. 46–47). Para uma objeção relacionada, confira Bueno e French (2011, p. 871–873). E apelar para fatos adicionais sobre o saleiro (o saleiro à direita do moinho de pimenta pode nos permitir inferir que Madagascar fica ao leste de Moçambique) para responder a essa objeção vai além do *Decreto Estipulativo*. Callender e Cohen admitem que algumas representações são mais úteis que outras, mas afirmam que

> as perguntas sobre a utilidade desses veículos representacionais são questões sobre a pragmática de coisas que são veículos representacionais, e não questões sobre seu *status* representacional *per se* (2006, p. 75).

Mas, mesmo que a *Condição de Raciocínio Substitutivo* seja relegada ao campo da "pragmática", parece razoável pedir uma explicação de como ela é atendida.

Uma coisa importante a ser observada é que, mesmo que o *Decreto Estipulativo* seja insustentável, não precisamos desistir do GG. O GG exige apenas que exista **alguma** explicação de como as representações derivadas se relacionam com as representações fundamentais; não exige que essa explicação seja de um tipo específico, muito menos que ela consista em nada além de um ato de estipulação (TOON, 2010, p. 77-78). Como Callender e Cohen observam, tudo o que é necessário é que exista uma classe privilegiada de representações e que outros tipos de representações devam suas capacidades representacionais à sua relação com as representações primitivas. Portanto, os filósofos precisam de uma explicação de como os membros dessa classe privilegiada de representações representam e como as representações derivadas, que incluem modelos científicos, se relacionam com essa classe. Quando afirmado assim, muitos colaboradores recentes do debate sobre representação científica podem ser vistos como estando sob a égide do GG.

De fato, como veremos abaixo, muitas das versões mais desenvolvidas das explicações de representação científica discutidas ao longo deste texto invocam as intenções dos usuários do modelo, embora de uma maneira mais complexa que o *Decreto Estipulativo*.

3. A Concepção de Similaridade

Similaridade (*similarity*) e representação inicialmente parecem ser dois conceitos intimamente relacionados, e invocar o primeiro para fundamentar o segundo tem uma linha filosófica que remonta, pelo menos, até *A República*, de Platão.[60] Em seu aspecto mais básico, a concepção de similaridade da representação científica afirma que os modelos científicos representam seus alvos em virtude de serem semelhantes a eles. Essa concepção tem aspirações universais, na medida em que é tomada para explicar a representação através de uma ampla gama de domínios diferentes. Dizem que pinturas, estátuas e desenhos são semelhantes aos seus modelos originais, e Giere proclamou que isso abrange modelos científicos ao lado de "palavras, equações, diagramas, gráficos, fotografias e, cada vez mais, imagens geradas por computador" (2004, p. 243). Portanto, a visão de similaridade repudia o problema da demarcação e sustenta que o mesmo mecanismo, a semelhança, sustenta diferentes tipos de representação em uma ampla variedade de contextos.

A visão também oferece uma elegante explicação do raciocínio substitutivo. Similaridades entre o modelo e o alvo podem ser exploradas para transportar as informações adquiridas no modelo para o alvo. Se a semelhança entre **M** e **T** for baseada em propriedades compartilhadas, então uma propriedade encontrada em **M** também deverá estar presente em **T**; e se a semelhança se mantiver entre as próprias propriedades, então **T** terá que instanciar propriedades similares a **M**.

No entanto, o apelo à semelhança no contexto da representação deixa em aberto se a similaridade é oferecida como resposta ao *Problema RE*, o *Problema do Estilo* ou se pretende definir *Padrões de Precisão*. Os proponentes da concepção

---

[60] Em estética, o termo *resemblance* é mais comumente usado do que *similarity*, mas não parece haver uma diferença substantiva entre as noções e as usamos de forma intercambiável. Confira Abell (2009) e Lopes (2004) para discussões relativamente atuais sobre similaridade na estética.

de similaridade geralmente oferecem pouca orientação sobre essa questão. Portanto, examinaremos cada opção e perguntaremos se a similaridade oferece uma resposta viável. Passaremos, então, à questão de como a visão de similaridade lida com o *Problema da Ontologia*.

3.1 Similaridade e Problema RE

Entendida como resposta ao Problema RE, a visão de similaridade mais simples é a seguinte:

> **Similaridade 1***:* Um modelo científico **M** representa um alvo **T** sse **M** e **T** são similares.

Uma objeção bem conhecida a essa explicação é que a similaridade tem as propriedades lógicas incorretas. Goodman (1976, p. 4-5) aponta que a similaridade é simétrica e reflexiva, mas a representação não. Se o objeto **A** é semelhante ao objeto **B**, então **B** é semelhante a **A**. Mas se **A** representa **B**, então **B** não precisa (e, de fato, na maioria dos casos não) representar **A**. Tudo é semelhante a si mesmo, mas a maioria das coisas não representam a si mesmas. Portanto, essa explanação não atende à nossa quarta condição de adequação para uma explicação da representação científica, na medida em que não fornece uma direção para a representação.

No entanto, existem relatos de similaridade sob os quais a similaridade não é uma relação simétrica (*vide* TVERSKY, 1977; WEISBERG, 2012, 2013, cap. 8; POZNIC, 2016, seção 4.2). Isso levanta a questão de como analisar a similaridade. Passaremos a esse problema na próxima subseção. No entanto, mesmo se admitirmos que a similaridade nem sempre precisa ser simétrica, isso não resolve o problema de Goodman com a reflexividade; nem, como veremos, resolve outros problemas da explicação da similaridade.

O problema mais significativo enfrentado pela **Similaridade 1** é que, sem restrições sobre o que conta como similar, quaisquer duas coisas podem ser consideradas similares (ARONSON *et al*. 1995, p. 21; GOODMAN, 1972, p. 443-444). Isso, no entanto, tem a consequência infeliz de que qualquer coisa representa qualquer outra coisa. Uma resposta natural a essa dificuldade é delinear um conjunto de aspectos e graus relevantes aos quais **M** e **T** devem ser similares. Essa ideia pode ser moldada na seguinte definição:

> **Similaridade 2:** Um modelo científico **M** representa um alvo **T** sse **M** e **T** forem similares em aspectos relevantes e em graus relevantes.

Nesta definição, é possível escolher os aspectos e graus de cada um, para que semelhanças indesejadas sejam excluídas. Enquanto isso resolve o último problema, deixa intocado o problema das propriedades lógicas: a semelhança em aspectos relevantes e em graus relevantes é reflexiva (e simétrica, dependendo da noção de semelhança). Além disso, a **Similaridade 2** enfrenta mais três problemas.

Em primeiro lugar, a similaridade, mesmo restrita a similaridades relevantes, é um conceito muito inclusivo para explicar a representação. Em muitos casos, nenhum par de objetos semelhantes representa o outro. Esse ponto foi trazido à tona em um clássico experimento de pensamento realizado por Putnam (1981, p. 1-3). Uma formiga está rastejando em um pedaço de areia e deixa um rastro que parece com Winston Churchill. A formiga produziu uma imagem, uma representação, de Churchill? A resposta de Putnam é que não, porque a formiga nunca viu Churchill, não tinha a intenção de produzir uma imagem dele, não estava causalmente ligada a Churchill, e assim por diante. Embora outra pessoa possa ver o traço como uma representação de Churchill, o traço em si não representa Churchill. O fato de o traço ser semelhante a Churchill não é suficiente para estabelecer que o traço o representa. E o que é verdade sobre o traço e Churchill é verdade sobre todos os outros pares de itens semelhantes: mesmo a similaridade relevante por si só não estabelece a representação.

Em segundo lugar, como observado na Seção 1, permitir a possibilidade de deturpação é um *desiderata* essencial exigido em qualquer explicação da representação científica. No contexto de uma concepção de similaridade, parece que uma deturpação é aquela que retrata seu alvo como tendo propriedades que não são semelhantes nos aspectos relevantes e, em grau relevante, às propriedades reais do alvo. Mas então, na **Similaridade 2**, **M** não é uma representação. A descrição, portanto, tem dificuldade em distinguir entre deturpação e não-representação (SUÁREZ, 2003, p. 233-235).

Terceiro, pode simplesmente não haver nada similar, porque algumas representações não representam um objeto real. Algumas pinturas representam elfos e dragões, e alguns modelos representam o flogisto e o éter. Nenhum deles existe. Esse é um problema para a visão de similaridade, porque os modelos sem alvos não podem representar o que parecem representar, porque, para que duas

coisas sejam semelhantes entre si, ambas precisam existir. Se não houver éter, então um modelo de éter não poderá ser similar ao éter.

Pelo menos alguns desses problemas podem ser resolvidos adotando o próprio ato de **afirmar** uma similaridade específica entre um modelo e um alvo como constitutivo da representação científica. Giere (1988, p. 81) sugere que os modelos são equipados com o que ele chama de "hipóteses teóricas", enunciados que afirmam que modelo e alvo são similares em aspectos relevantes e em certos graus. Ele enfatiza que "os cientistas são agentes intencionais com objetivos e propósitos" (2004, p. 743) e propõe incorporar essa percepção explicitamente em uma explicação da representação. Isso envolve a adoção de uma noção de representação baseada em agentes que se concentra na "atividade de representar" (2004). Analisar a representação nesses termos equivale a analisar esquemas como

> Os agentes (1) têm a intenção de; (2) usar o modelo *M*; (3) para representar uma parte do mundo *W*; (4) para propósitos *P*. Portanto, os agentes especificam quais similaridades são pretendidas e com qual finalidade (2010, p. 274). Confira também Mäki (2009, 2011). E, para argumentos contrários, veja Rusanen e Lappi (2012, p. 317).

Isso leva à seguinte definição:

> **Similaridade 3:** Um modelo científico **M** representa um sistema-alvo **T** sse houver um agente **A** que use **M** para representar um sistema-alvo **T**, propondo uma hipótese teórica **H**, especificando uma similaridade (em certos aspectos e em certos graus) entre **M** e **T**, para a finalidade **P**.

Essa versão da visão de similaridade evita problemas de deturpação porque **H** é uma hipótese, e assim não há expectativa de que as afirmações feitas em **H** sejam verdadeiras. Se elas são, então a representação é precisa (ou a representação é precisa na medida em que elas se sustentam). Caso contrário, a representação é uma deturpação. Também resolve o problema com a direcionalidade, porque **H** pode ser entendido como um fator que introduz uma assimetria que não está presente na relação de similaridade. No entanto, ela não resolve o problema com a representação

sem um alvo. Se não houver éter, nenhuma hipótese poderá ser afirmada a respeito, pelo menos de maneira direta.

A **Similaridade 3**, ao invocar um papel ativo para os propósitos e ações dos cientistas na constituição da representação científica, marca uma mudança significativa na ênfase das explicações baseadas em similaridades. Suárez (2003, pp. 226–227), com base em van Fraassen (2002) e Putnam (2002), define explicações "naturalistas" sobre a representação como aquelas em que

> a representação obtida depende ou não de fatos sobre o mundo e de forma alguma responde aos propósitos pessoais, pontos de vista ou interesses dos investigadores.

Por envolver diretamente os propósitos dos usuários do modelo em resposta ao problema RE, a **Similaridade 3** não é explicitamente uma explicação naturalista (em contraste à **Similaridade 1**).

Embora a **Similaridade 3** resolva vários problemas que envolvem versões mais simples, ela não parece ser uma solução bem-sucedida, baseada em similaridade, para o Problema RE. Um exame mais detalhado da **Similaridade 3** revela que o papel da similaridade mudou. No que diz respeito à oferta de uma solução para o problema RE, todo o trabalho pesado na **Similaridade 3** é feito pelo apelo aos agentes e suas intenções. Giere admite isso implicitamente quando observa que a similaridade é "a maneira mais importante, mas provavelmente não é a única" para os modelos funcionarem representacionalmente (2004, p. 747). Mas se a similaridade não é a única maneira pela qual um modelo pode ser usado como uma representação, a similaridade se torna inútil em resposta ao problema RE. De fato, ser similar nos aspectos relevantes ao grau relevante agora desempenha o papel de um estilo representacional ou de um critério normativo para uma representação precisa, em vez de constituir uma representação *per se*. Avaliaremos na próxima seção se a similaridade oferece uma resposta convincente às questões de estilo e precisão e levantaremos um problema adicional para qualquer descrição da representação científica que se baseie na ideia de que modelos, especificamente modelos não concretos, são similares aos seus alvos.

3.2 Precisão, Estilo e Ontologia

O fato de que propriedades relevantes podem ser delineadas de maneiras diferentes pode potencialmente fornecer uma resposta ao Problema do Estilo. Se o fato de que **M** representa **T** envolve a alegação de que **M** e **T** são similares em um certo aspecto, o aspecto escolhido especifica o estilo da representação; e se **M** e **T** são de fato semelhantes neste aspecto (e no grau especificado), então **M** representa com precisão **T** dentro desse estilo. Por exemplo, se é sugerido que **M** e **T** são similares em relação à sua estrutura causal, então podemos ter um estilo de modelagem causal; se é sugerido que **M** e **T** são similares em relação às propriedades estruturais, então podemos ter um estilo de modelagem estrutural; e assim por diante.

Um primeiro passo na direção de tal compreensão de estilos é a análise explícita da noção de similaridade. A maneira padrão de entender o que significa para um objeto ser semelhante a outro objeto é exigindo que eles co-instanciem propriedades. De fato, essa é a ideia que Quine (1969, p. 117-118) e Goodman (1972, p. 443) tinham em mente em suas influentes críticas à similaridade. As duas estruturas formais mais importantes que desenvolvem essa ideia são as explicações geométrica e de contraste. Para uma discussão a respeito, confira Decock e Douven (2011).

A explicação geométrica, associada a Shepard (1980), atribui aos objetos um lugar em um espaço multidimensional com base nos valores atribuídos às suas propriedades. Esse espaço é, então, equipado com uma métrica e o grau de (des)similaridade entre os dois objetos é a distância entre os pontos que representam os dois objetos naquele espaço. Essa explicação baseia-se nas fortes suposições de que os valores podem ser atribuídos a todas as características relevantes para os julgamentos de similaridade, que são considerados irrealistas (e, até onde sabemos, ninguém desenvolveu tal explicação no contexto da representação científica).

Supõe-se que este problema é superado na explicação de contraste de Tversky (1977). Esta explicação define uma noção gradual de similaridade com base em uma comparação ponderada de propriedades. Recentemente, Weisberg introduziu essa explicação na filosofia da ciência, onde serve como ponto de partida para sua **correspondente explicação das características ponderadas do modelo de relações com o mundo** (*vide* WEISBERG, 2012, 2013, cap. 8). Embora a explicação tenha algumas vantagens, permanecem questões sobre se ela pode capturar a distinção entre o que Niiniluoto (1988, p. 272-274) chama de "similitude" (*likeness*) e "identidade parcial". Dois objetos são parecidos na medida

em que co-instanciam propriedades semelhantes (por exemplo, uma cabine telefônica vermelha e um ônibus vermelho de Londres podem ser iguais em relação à cor, apesar de não instanciarem exatamente o mesmo tom de vermelho). Dois objetos são parcialmente idênticos na medida em que co-instanciam propriedades idênticas. Como Parker (2015, p. 273) observa, explicações de similaridade baseadas em contraste como os de Weisberg têm dificuldades em capturar o primeiro, e isso é frequentemente pertinente no contexto da representação científica em que os modelos e seus alvos não precisam co-instanciar exatamente a mesma propriedade.

Uma outra pergunta que permanece para alguém que usa a noção de similaridade para responder ao Problema do Estilo e providenciar padrões de precisão da maneira aqui considerada é se ele realmente captura toda a prática científica. Os teóricos da similaridade estão comprometidos com a alegação de que sempre que um modelo científico representa seu sistema-alvo, isso é estabelecido em virtude de um usuário do modelo especificar uma similaridade relevante e, se a similaridade se mantiver, então a relação representacional é precisa. Essas aspirações universais exigem que a noção de similaridade invocada capture a relação que é mantida entre as diversas entidades, como a máquina Phillips-Newlyn e uma economia, um mapa de tubos e um sistema de trens subterrâneos, as equações de Lotka-Volterra (ou o espaço de fase associado a elas) e populações de predadores e presas, entre outros. Se todas essas relações podem ser capturadas em termos de similaridade permanece uma questão em aberto.

Outro problema enfrentado por abordagens baseadas em similaridade diz respeito ao tratamento da ontologia dos modelos. Se os modelos devem ser semelhantes aos seus alvos das maneiras especificadas pelas hipóteses teóricas, então eles devem ser o **tipo** de coisa que podem ser tão semelhantes. Para modelos materiais como o modelo da Baía de São Francisco (WEISBERG, 2013), modelos de moléculas de bola e bastão (TOON, 2011), a máquina Phillips-Newlyn (MORGAN; BOUMANS, 2004) ou organismos modelo (ANKENY; LEONELLI, 2011) isso parece simples porque eles são do mesmo tipo ontológico que seus respectivos alvos. Mas muitos modelos científicos interessantes não são assim: eles são o que Hacking descreve apropriadamente como "algo que você tem na cabeça e não nas mãos" (1983, p. 216). Seguindo Thomson-Jones (2012), chamamos esses modelos de **modelos não concretos**. A questão, então, é como esses modelos podem ser similares aos seus alvos. No mínimo, esses modelos são "abstratos" no sentido de que não têm localização espaço-temporal. Mas, nesse caso, ainda não está claro como eles podem instanciar os tipos de propriedades especificados por hipóteses teóricas,

uma vez que essas propriedades são tipicamente *físicas* e, presumivelmente, o fato estar localizadas no espaço e no tempo é uma condição necessária para instanciar tais propriedades. Para uma discussão mais aprofundada sobre essa objeção e soluções propostas, consulte Teller (2001, p. 399), Thomson-Jones (2010) e Giere (2009).

4. A concepção estruturalista

A concepção estruturalista da representação do modelo originou-se na chamada visão semântica das teorias que ganhou destaque na segunda metade do século XX . A visão semântica foi originalmente proposta como uma explicação da estrutura da teoria, e não da representação científica. A ideia motriz que está por trás da posição é que as teorias científicas são melhor consideradas como coleções de modelos. Isso convida às perguntas: quais são esses modelos e como eles representam seus sistemas-alvo? A maioria dos defensores da visão semântica das teorias (com a exceção notável de Giere, cujas visões sobre representação científica foram discutidas na seção anterior) considera os modelos como estruturas, que representam seus sistemas-alvo em virtude de haver algum tipo de **morfismo** (isomorfismo, isomorfismo parcial, homomorfismo, etc.) entre os dois.

Essa concepção tem duas vantagens *prima facie*. A primeira vantagem é que ela oferece uma resposta direta ao Problema RE (ou problema RC, se o foco estiver na representação científica), e explica o raciocínio substitutivo: os mapeamentos entre o modelo e o alvo permitem que os cientistas convertam as verdades encontradas no modelo em afirmações sobre o sistema-alvo. A segunda vantagem diz respeito à aplicabilidade da matemática. Há uma posição consagrada na filosofia da matemática que vê a matemática como o estudo de estruturas; ver, por exemplo, Resnik (1997) e Shapiro (2000). É natural que o estruturalista científico adote esse ponto de vista, que fornece uma explicação clara de como a matemática é usada na modelagem científica.

4.1 Estruturas e o problema da ontologia

Quase tudo, desde uma sala de concertos a um sistema de parentesco, pode ser referido como uma "estrutura". Portanto, a primeira tarefa para uma explicação estruturalista da representação é articular qual noção de estrutura ela emprega. Várias noções diferentes de estrutura foram discutidas na literatura

(THOMSON-JONES, 2011), mas, de longe, a mais comum é a noção de estrutura que se encontra na teoria dos conjuntos e na lógica matemática. Uma estrutura **S** nesse sentido (às vezes "estrutura matemática" ou "estrutura teórica dos conjuntos") é uma entidade composta que consiste no seguinte: um conjunto não vazio de objetos **U** chamado domínio (ou universo) da estrutura e um conjunto **R** indexado de relações em **U** (apoiadores da abordagem de estruturas parciais, por exemplo, Da Costa e French (2003) e Bueno, French e Ladyman (2002), usam relações parciais de n-lugares, para as quais pode ser indefinido se algumas n-tuplas devem ou não estar em sua extensão). Essa definição de estrutura é amplamente usada em matemática e lógica. Observamos, no entanto, que na lógica matemática as estruturas também contêm uma linguagem e uma função interpretativa, interpretando símbolos da linguagem em termos de **U** (MACHOVER, 1996; HODGES, 1997), que está ausente das estruturas no contexto atual. É conveniente escrevê-las como **S** = <**U**, **R**>, onde "<,>" indica uma tupla ordenada.

É importante esclarecer o que queremos dizer com "objeto" e "relação" neste contexto. No que diz respeito aos objetos, tudo o que importa do ponto de vista estruturalista é que existem muitos e muitos deles. Se os objetos são mesas ou planetas é irrelevante. Tudo o que precisamos são manequins ou espaços reservados, cuja única propriedade é a "objetidade" [*objecthood*]. Da mesma forma, ao definir relações, desconsideramos completamente o que a relação é "em si". Quer falemos de "ser mãe de" ou "ficar à esquerda de" não é de interesse no contexto de uma estrutura; tudo o que importa é entre quais objetos se dar a relação. Por esse motivo, uma relação é especificada puramente de forma extensional: como uma classe de n-tuplas ordenadas. A relação literalmente não é nada além dessa classe. Portanto, uma estrutura consiste em objetos fictícios entre os quais se mantêm relações definidas de forma totalmente extensional.

A primeira posição básica da teoria estruturalista da representação é que os modelos são estruturas nesse sentido (a segunda é que os modelos representam seus alvos por serem adequadamente mórficos a eles; discutiremos morfismos na próxima subseção). Suppes articulou claramente essa posição quando declarou que "o significado do conceito de modelo é o mesmo na matemática e nas ciências empíricas" (1969, p. 12), e muitos o seguiram. Então, somos apresentados a uma resposta clara ao Problema da Ontologia: modelos são estruturas. A questão remanescente é o que são as próprias estruturas. São entidades Platônicas, classes de equivalência, construções modais ou ainda outra coisa? No contexto de uma discussão sobre representação científica, podemos empurrar essas questões para a filosofia da matemática.

4.2 Estruturalismo e o Problema RE

A concepção estruturalista mais básica da representação científica afirma que os modelos científicos, entendidos como estruturas, representam seus sistemas-alvo em virtude de serem isomórficos a eles. Um isomorfismo entre duas estruturas **S** e **S'** é uma função bijetora de **U** a **U'** que preserva as relações em **U** (e inversamente, as relações em **U'**). Um isomorfismo associa cada objeto em **U** a um objeto em **U'** e emparelha cada relação em **R** com uma relação em **R'**, de modo que uma relação se mantém entre certos objetos em **U** sse a relação correspondente se mantiver entre os objetos em **U'** que estão associados com eles.[61] Suponha agora que o sistema-alvo **T** exiba a estrutura $\mathbf{S_T}$ e o modelo seja a estrutura $\mathbf{S_M}$. Então, o modelo representa o alvo sse ele for isomórfico ao alvo:

> **Estruturalismo 1**: Um modelo científico **M** representa seu alvo **T sse** $\mathbf{S_T}$ é isomórfico a $\mathbf{S_M}$.

É importante notar que poucos adeptos da explicação estruturalista da representação científica, mais intimamente associada à visão semântica das teorias, defendem explicitamente essa posição (UBBINK, 1960, p. 302). A representação não era o foco de atenção na visão semântica, e a atribuição de (algo como) o **Estruturalismo 1** a seus apoiadores é uma extrapolação. A representação se tornou um tópico muito debatido na primeira década do século XXI, e muitos proponentes da visão semântica ou se afastaram do **Estruturalismo 1** ou apontaram que nunca sustentaram tal visão. Em breve, voltaremos para posições mais avançadas, mas para entender o que motiva essas posições, é útil entender por que o **Estruturalismo 1** falha.

O primeiro e mais óbvio problema é o mesmo da visão de similaridade: o isomorfismo é simétrico e reflexivo (e transitivo), mas a representação não é. Esse problema pode ser solucionado substituindo o isomorfismo por um mapeamento alternativo. Bartels (2006), Lloyd (1984) e Mundy (1986) sugerem homomorfismo; van Fraassen (1980, 1997, 2008) e Redhead (2001) inserções isomórficas; os defensores da abordagem de estruturas parciais preferem isomorfismos parciais (BUENO, 1997; BUENO; FRENCH, 2011; DA COSTA; FRENCH, 1990, 2003;

---

[61] Para saber mais sobre as dificuldades de resgatar essa noção de correspondência sem referência a uma função de interpretação, confira Halvorson (2012) e Glymour (2013).

FRENCH, 2003, 2014; FRENCH; LADYMAN, 1999); e Swoyer (1991) introduz o que ele chama de morfismos Δ / Ψ. Nós nos referimos a estes coletivamente como "morfismos".

Essas sugestões resolvem alguns, mas não todos os problemas. Embora muitos desses mapeamentos não sejam simétricos, eles ainda são reflexivos. Mas, mesmo que essas questões formais pudessem ser resolvidas de uma maneira ou de outra, uma visão baseada em mapeamentos estruturais ainda enfrentaria outros problemas sérios. Para facilitar a apresentação, discutiremos esses problemas no contexto da visão do isomorfismo; *mutatis mutandis* outros mapeamentos formais sofrem das mesmas dificuldades (para discussões detalhadas sobre homomorfismo e isomorfismo parcial, ver Suárez (2003, pp. 239-241) e Pero e Suárez (2016); Mundy (1986) discute restrições gerais que se pode impor aos morfismos). Como a similaridade, o isomorfismo é muito inclusivo: nem todas as coisas que são isomórficas se representam. No caso da similaridade, essa questão foi trazida à tona pelo experimento mental de Putnam com a formiga rastejando na praia; no caso do isomorfismo, uma olhada na história da ciência fará esse trabalho. Muitas estruturas matemáticas foram descobertas e discutidas muito antes de serem usadas na ciência. Geometrias não Euclidianas foram estudadas por matemáticos muito antes de Einstein utilizá-las no contexto das teorias do espaço-tempo, e os espaços de Hilbert foram estudados por matemáticos antes de serem utilizados na teoria quântica. Se a representação não fosse nada além do isomorfismo, teríamos que concluir que Riemann descobriu a relatividade geral ou que Hilbert inventou a mecânica quântica. Isso não parece correto, portanto, não parece que o isomorfismo por si só estabeleça uma representação científica (FRIGG, 2002, p. 10).

O isomorfismo é mais restritivo que a similaridade: nem tudo é isomórfico a todas as outras coisas. Mas o isomorfismo ainda é abundante demais para identificar corretamente o que um modelo representa. A raiz das dificuldades é que as mesmas estruturas podem ser instanciadas em diferentes tipos de sistemas-alvo. Certas estruturas geométricas são instanciadas por muitos sistemas diferentes; pense em quantas coisas esféricas encontramos no mundo. A lei $1/\mathbf{r}^2$ da gravidade Newtoniana também é o "esqueleto matemático" da lei de atração eletrostática de Coulomb e o enfraquecimento do som ou da luz em função da distância da fonte. A estrutura matemática do pêndulo também é a estrutura de um circuito elétrico com um condensador e um solenóide (KROES, 1989). A mesma estrutura pode ser exibida por mais de um tipo de sistema-alvo e, portanto, o isomorfismo por si só é muito fraco para identificar o alvo de um modelo.

Como vimos na última seção, uma deturpação é aquela que retrata seu alvo como tendo características que não possui. No caso de uma explicação estrutural da representação, isso significa que o modelo retrata o alvo como tendo propriedades estruturais que ele não possui. No entanto, o isomorfismo exige identidade da estrutura: as propriedades estruturais do modelo e do alvo devem corresponder exatamente uma a outra. Portanto, uma deturpação não será isomórfica ao alvo. À luz do **Estruturalismo 1**, não é, portanto, uma representação em absoluto. Como simples descrições de similaridade, o **Estruturalismo 1** confunde deturpação com não representação (SUÁREZ, 2003, p. 234-235). Estruturas parciais podem evitar uma incompatibilidade devido a relação-alvo ser omitida no modelo e, portanto, de alguma forma reforçam a explicação estruturalista (BUENO; FRENCH 2011, p. 888). Ainda não está claro, entretanto, como eles explicam as representações distorcidas (PINCOCK, 2005).

Por fim, como as explicações de similaridade, o **Estruturalismo 1** tem um problema com alvos inexistentes porque nenhum modelo pode ser isomórfico a algo que não existe. Se não houver éter, um modelo não pode ser isomórfico a ele. Portanto, os modelos sem alvo não podem representar o que parecem representar.

A maioria desses problemas pode ser resolvido fazendo movimentos semelhantes aos que levam à **Similaridade 3**: introduzir agentes e raciocínio hipotético na explicação da representação. Analisando os movimentos, encontramos:

> **Estruturalismo 2**: Um modelo científico **M** representa um sistema-alvo **T sse** houver um agente **A** que use **M** para representar um sistema-alvo **T**, propondo uma hipótese teórica **H** e especificando um isomorfismo entre $\mathbf{S_M}$ e $\mathbf{S_T}$.

Isso está de acordo com os recentes pronunciamentos de van Fraassen sobre representação. Ele oferece a seguinte descrição como o "Hauptstatz" de uma teoria da representação: "*Não há representação, exceto no sentido de que algumas coisas são usadas, fabricadas ou tomadas para representar as coisas como isto ou aquilo*" (2008, p. 23, grifo do autor). Da mesma forma, Bueno afirma que "representação é um ato *intencional* que relaciona dois objetos" (2010, p. 94-95, grifo do autor), e Bueno e French apontam que usar uma coisa para representar outra coisa não é apenas uma função de isomorfismo (parcial) mas também depende de fatores “pragmáticos” “relacionados com o uso que damos aos modelos

relevantes" (2011, p. 885).

Como na mudança de **Similaridade 2** para **Similaridade 3**, esta parece uma mudança bem-sucedida, com muitas (embora não todas) das preocupações mencionadas acima sendo atendidas. Mas, novamente, o papel do isomorfismo mudou. O ingrediente crucial é a intenção do agente e o isomorfismo tornou-se de fato ou um estilo representacional ou critério normativo para uma representação precisa. Vamos agora avaliar quão bem se sai o isomorfismo como uma resposta a esses problemas e aos outros descritos acima.

4.3 Demarcação, Precisão, Estilo e Estrutura Alvo-final

A posição do estruturalismo sobre o Problema da Demarcação é, em geral, uma questão em aberto. Ao contrário da similaridade, que tem sido amplamente discutida em diferentes domínios, os mapeamentos estruturais estão intimamente ligados à estrutura formal da teoria dos conjuntos e têm sido discutidos apenas com moderação fora do contexto das ciências matematizadas. Uma exceção é French (2003), que discute explicações de isomorfismo no contexto da representação pictórica. Ele discute em detalhes a explicação da representação pictórica de Budd (1993) e aponta que é baseado na noção de um isomorfismo estrutural entre a estrutura da superfície da pintura e a estrutura do campo visual relevante. Portanto, a representação é o isomorfismo percebido da estrutura (FRENCH, 2003, p. 1475-1476), reafirmado por Bueno e French (2011, p. 864-865), e para o qual há uma discussão crítica de Downes (2009, p. 423-425).

O Problema do Estilo é identificar estilos representacionais e caracterizá-los. Um mapeamento estrutural proposto entre o modelo e o alvo oferece uma resposta óbvia a este desafio: pode-se representar um sistema criando um modelo que se propõe a ser apropriadamente mórfico a ele. Isso fornece o estilo do isomorfismo, o estilo do homomorfismo, o estilo do isomorfismo parcial e assim por diante. Podemos chamá-los de "estilos de morfismo" quando nos referimos a eles em geral. Cada um desses estilos também oferece uma condição bem definida de precisão: a representação é precisa se o morfismo hipotético se mantém; é impreciso se não o fizer.

Esta é uma resposta clara. A questão é que status ele tem em relação ao Problema do Estilo. Os estilos de morfismo são apenas um subgrupo de estilos ou são privilegiados? O primeiro é incontestável. No entanto, a ênfase que muitos

estruturalistas colocam nos mapeamentos de preservação de estrutura sugere que eles não consideram os morfismos apenas como uma forma entre outras de representar algo. O que eles parecem ter em mente é a afirmação mais forte de que uma representação *deve* ser desse tipo, ou que os estilos de morfismo são os únicos estilos aceitáveis.

Essa afirmação parece conflitar com a prática científica em pelo menos dois aspectos. Em primeiro lugar, muitas representações são imprecisas (e conhecidas por serem) de alguma forma. Alguns modelos distorcem, deformam e torcem as propriedades do alvo de forma que parecem reduzir o isomorfismo, ou mesmo qualquer um dos mapeamentos de preservação da estrutura proposta. Alguns modelos em mecânica estatística têm um número infinito de partículas e o modelo Newtoniano do sistema solar representa o sol como uma esfera perfeita, onde na realidade é uma bola de fogo sem nenhuma superfície bem definida. Na melhor das hipóteses, não está claro como o isomorfismo, parcial ou não, ou homomorfismo pode explicar esses tipos de idealizações. Portanto, parece que estilos de representação diferentes dos mapeamentos que preservam a estrutura devem ser reconhecidos.

Em segundo lugar, a visão estruturalista é uma reconstrução racional da modelagem científica e, como tal, tem alguma distância da prática real. Alguns filósofos se preocupam que essa distância seja muito grande e que a visão esteja muito distante da prática real da ciência para poder captar o que importa para a prática da modelagem. Esse é o impulso de muitas contribuições para Morgan e Morrison (1999), confira, também, Cartwright (1999). Embora alguns modelos usados pelos cientistas possam ser melhor considerados como estruturas teóricas definidas, há muitos casos em que isso parece contradizer como os cientistas realmente falam e argumentam sobre seus modelos. Exemplos óbvios incluem modelos físicos como o modelo da Baía de São Francisco (WEISBERG, 2013), mas também sistemas como o pêndulo idealizado ou populações imaginárias de animais intercruzados. Tais modelos têm a estranha propriedade de serem concretos se reais [*concrete-if-real*] e os cientistas falam sobre eles como se fossem sistemas reais, apesar do fato de que obviamente não são (GODFREY-SMITH, 2006). Thomson-Jones (2010) chama isso de "prática de valor de face", e há uma questão de se o estruturalismo pode ser responsável por essa prática.

Resta um último problema a ser abordado no contexto das explicações estruturais da representação científica. Os sistemas-alvo são objetos físicos: átomos, planetas, populações de coelhos, agentes econômicos etc. O isomorfismo é uma relação que existe entre duas estruturas e alegar que uma estrutura teórica definida

é isomórfica a um pedaço do mundo físico é prima facie um erro de categoria. Por definição, um morfismo só pode ocorrer entre duas estruturas. Se quisermos dar sentido à afirmação de que o modelo é isomórfico ao seu alvo, temos que assumir que o alvo de alguma forma exibe uma certa estrutura $\mathbf{S}_T$. Mas o que significa para um sistema-alvo – uma parte do mundo físico – possuir uma estrutura, e onde no sistema-alvo a estrutura está localizada?

Existem duas sugestões proeminentes na literatura. A primeira, originalmente sugerida por Suppes (1969), é que os modelos de dados são as estruturas alvo-final [*target-end*] representadas por modelos. Esta abordagem enfrenta a questão de saber se devemos ficar satisfeitos com uma explicação da representação científica que impede que os fenômenos sejam representados (ver Bogen e Woodward (1988) para uma discussão sobre a distinção entre dados e fenômenos, e Brading e Landry (2006) para uma discussão da distinção no contexto da representação científica). Van Fraassen (2008) abordou esse problema e defende uma resolução pragmática: no contexto do uso, não há diferença pragmática entre representar fenômenos e dados extraídos deles (ver Nguyen 2016 para uma discussão crítica). A abordagem alternativa localiza a estrutura alvo-final no próprio sistema-alvo. Uma versão dessa abordagem vê as estruturas como sendo **instanciadas** nos sistemas-alvo. Essa visão parece estar implícita em muitas versões da visão semântica e é explicitamente sustentada por autores que defendem uma resposta estruturalista para o problema da aplicabilidade da matemática (RESNIK, 1997; SHAPIRO, 1997). Essa abordagem enfrenta problemas de subdeterminação, pois o mesmo alvo pode instanciar estruturas diferentes. O problema pode ser visto como surgindo devido à existência de descrições alternativas do sistema (FRIGG, 2006) ou porque uma versão da "Objeção de Newman" também atinge o contexto atual (NEWMAN, 1928; AINSWORTH, 2009). Para uma discussão aprofundada, confira Ketland (2004). Uma versão mais radical simplesmente **identifica** alvos com estruturas (TEGMARK, 2008). Esta abordagem é altamente revisionária, em particular quando se considera sistemas-alvo como populações de coelhos reprodutores ou economias. Portanto, a questão permanece para qualquer explicação estruturalista da representação científica: onde as estruturas alvo-final requeridas podem ser encontradas?

5. A Concepção Inferencial

A ideia central da concepção inferencial é analisar a representação científica em termos da função inferencial dos modelos científicos. Nas explicações anteriormente discutidas, a capacidade inferencial de um modelo abandonou tudo aquilo que foi suposto para responder ao problema RE (ou Problema RC): morfismos propostos ou relações de similaridade entre os modelos e seus alvos, por exemplo. As explicações discutidas nesta seção invertem essa ordem e explicam a representação científica diretamente em termos de raciocínio substitutivo.

5.1 A Explicação DDI

De acordo com a explicação *Denotação, Demonstração, e Interpretação* (DDI) acerca da representação científica proposta por Hughes (1997, 2010, cap. 5), os modelos **denotam** seus alvos; são tais que os usuários do modelo podem realizar *demonstrações* neles; e **interpretar** os resultados de tais demonstrações em termos do alvo. A última etapa é necessária porque as demonstrações estabelecem resultados sobre o próprio modelo e, ao interpretar esses resultados, o usuário do modelo tira inferências sobre o alvo a partir do modelo (1997, p. 333). Infelizmente, Hughes tem pouco a dizer sobre o que significa interpretar um resultado de uma demonstração em um modelo em termos de seu sistema-alvo, e então é preciso recuar para uma noção intuitiva (e não analisada) de fazer inferências sobre o alvo com base no modelo.[62]

Hughes deixa claro que ele não está tentando responder ao problema RE, e que ele não oferece denotação, demonstração e interpretação como individualmente necessárias e nem conjuntamente como condições suficientes para a representação científica. Ele prefere

> a sugestão mais modesta de que, se examinarmos um modelo teórico com essas três atividades em

[62] a sugestão mais modesta de que, se examinarmos um modelo teórico com essas três atividades em mente, alcançaremos algum insight sobre o tipo de representação que elas fornecem (1997, p. 339).

> mente, alcançaremos algum insight sobre o tipo de representação que elas fornecem (1997, p. 339).

Isso é insatisfatório porque, em última análise, permanece obscuro o que permite aos cientistas usar um modelo para fazer inferências sobre o alvo, e levanta a questão do que teria de ser adicionado às condições DDI para transformá-las em uma resposta completa ao problema RE. Se, alternativamente, as condições forem consideradas necessárias e suficientes, então, a explicação exigiria uma elaboração mais detalhada sobre o que estabelece as condições.

5.2 Inferencialismo Deflacionário

Suárez argumenta que devemos adotar uma "atitude e estratégia deflacionária ou minimalista" (2004: 770) ao abordar o problema da representação científica. Duas noções diferentes de deflacionismo estão em operação na sua explicação. O primeiro é abandonar o objetivo de buscar as condições necessárias e suficientes; as condições necessárias serão boas o suficiente (2004, p. 771). A segunda noção é que não devemos buscar "nenhuma característica mais profunda da representação além de suas características superficiais" (2004, p. 771) ou "superficialidades" (Suárez e Solé 2006, p. 40), e que devemos negar que uma análise de um conceito " é o tipo de análise que lançará luz explicativa sobre nosso uso do conceito "(Suárez 2015, p. 39). Suárez pretende que sua explicação da representação científica seja deflacionária em ambos os sentidos, e o denomina de "inferencialismo". Deixando **A** representar o modelo e **B** o alvo, ele oferece a seguinte análise:

> **Inferencialismo:** "**A** representa **B** apenas se (i) a força representacional de **A** aponta para **B**, e (ii) **A** permite que agentes competentes e informados façam inferências específicas sobre **B**" (2004, p. 773).

A primeira condição aborda o Requisito de Direcionalidade e garante que **A** e **B** realmente estabeleçam uma relação de representação. Poderia surgir a preocupação com o fato de que explicar a representação em termos de força representacional lança pouca luz sobre o assunto, se nenhuma análise dessa força for oferecida. Mas Suárez resiste às tentativas de explicar a força representacional

em termos de uma relação mais forte, como denotação ou referência, sob o argumento de que isso violaria o deflacionismo (2015, p. 41). A segunda condição é, na verdade, apenas a Condição de Raciocínio Substitutivo, agora considerada uma condição necessária na representação científica. Mas Contessa (2007, p. 61) aponta que permanece um mistério como essas inferências são geradas. Um apelo a uma análise mais aprofundada pode, novamente, ser bloqueado pelo apelo ao deflacionismo, porque qualquer tentativa de explicar como as inferências são traçadas iria além das "características superficiais". Portanto, a sustentabilidade do **Inferencialismo** em vigor depende da sustentabilidade do deflacionismo sobre a representação científica. Suárez (2015) defende o deflacionismo traçando analogias com três diferentes teorias deflacionárias sobre a verdade, a teoria da "redundância" de Ramsey, o "minimalismo abstrato" de Wright e a "teoria do uso" de Horwich. Uma defesa alternativa baseia-se no inferencialismo de Brandom na filosofia da linguagem (1994, 2000), uma linha de argumento que é desenvolvida em de Donato Rodríguez e Zamora Bonilla (2009).

O **Inferencialismo** fornece uma explicação clara da possibilidade de deturpação, porque as inferências feitas sobre um alvo não precisam ser verdadeiras (Suárez 2004, p. 776). Na medida em que se aceita a força representacional como um conceito convincente, os modelos sem alvo são tratados com sucesso porque a força representacional (diferentemente da denotação) não requer a existência de um objetivo (2004, p. 772). O inferencialismo repudia o Problema da Demarcação Representacional e visa oferecer uma explicação da representação que também funciona em outros domínios, como a pintura (2004, p. 777). A explicação está ontologicamente não comprometida, porque qualquer coisa que tenha uma estrutura interna que permita ao agente fazer inferências pode ser uma representação. De modo semelhante, como a explicação deve ser aplicada a uma ampla variedade de entidades, incluindo equações e estruturas matemáticas, ela implica que a matemática é aplicada com êxito nas ciências, mas, de acordo com o espírito do deflacionismo, não há explicação sobre como isso é possível. A explicação não aborda diretamente o Problema do Estilo.

5.3 Inflando o Inferencialismo: Interpretação

Em resposta às dificuldades com o **Inferencialismo**, Contessa afirma que "não está claro por que devemos adotar uma atitude deflacionária desde o i*nício*" (2007,

p. 50) e fornece uma "explicação interpretativa" da representação científica que é inspirada pela explicação de Suárez, mas sem ser deflacionária. Contessa introduz a noção de uma interpretação de um modelo, em termos de seus sistemas-alvo, como uma condição necessária e suficiente na representação epistêmica. (DUCHEYNE, 2012):

> **Interpretação:** "Um [modelo **M**] é uma representação epistêmica de um determinado alvo [**T**] (para um determinado usuário) se e somente se o usuário adotar uma interpretação de [**M**] em termos de [**T**]" (Contessa 2007, p. 57; 2011, p. 126-127).

A ideia principal de uma interpretação é que o usuário do modelo primeiro identifique os conjuntos de objetos relevantes no modelo e no alvo e, em seguida, fixe os conjuntos de propriedades e relações que esses objetos instanciam no modelo e no alvo. O usuário então (a) toma **M** para denotar **T**; (b) considera cada objeto identificado no modelo para denotar exatamente um objeto no alvo (e cada objeto relevante no alvo tem que ser denotado); (c) considera cada propriedade e relação no modelo para denotar uma propriedade ou relação do mesmo tipo no alvo (e, novamente, cada propriedade e relação no alvo tem que ser denotada). Uma tradução formal dessas condições é o que Contessa chama de "interpretação analítica" (2007, p. 57-62). Ele também inclui uma condição adicional referente às funções no modelo e no alvo, que suprimimos para sermos mais breves.

A **Interpretação** oferece uma resposta simples para o problema RE. Ela também explica a direcionalidade da representação: interpretar um modelo em termos de um alvo não implica interpretar um alvo em termos de um modelo. No entanto, foi observado que a **Interpretação** tem dificuldade em explicar a possibilidade de deturpação, uma vez que parece exigir que os objetos, propriedades e relações relevantes realmente existam no alvo (SHECH, 2015), embora esta objeção seja uma leitura muito estrita da explicação de Contessa. Contessa não comenta sobre a aplicabilidade da matemática, mas uma vez que sua explicação compartilha com a explicação estruturalista uma ênfase nas relações e na correspondência modelo-alvo um para um, Contessa pode apelar para a mesma descrição da aplicabilidade da matemática realizada pelo estruturalista. Como Suárez, Contessa considera sua explicação universal e aplicável a representações não científicas, como retratos e mapas. Mas ainda não está claro como a **Interpretação** aborda o Problema do Estilo. Como vimos anteriormente, em particular as representações visuais se

enquadram em diferentes categorias. É uma questão para pesquisas futuras como elas podem ser classificadas dentro da estrutura interpretativa. No que diz respeito à Questão da Ontologia, a própria **Interpretação** impõe poucas restrições ao que são os modelos científicos, ontologicamente falando. Tudo o que se requer é que eles consistam em objetos, propriedades, relações e funções. Confira Contessa (2010) para uma discussão mais aprofundada sobre sua posição sobre os modelos, ontologicamente falando.

6. A Visão Ficcional dos Modelos

Uma família recente de abordagens analisa modelos traçando uma analogia entre modelos e ficção literária. Essa analogia pode ser usada de duas maneiras, resultando em duas diferentes versões da visão ficcional. A primeira é motivada principalmente por considerações ontológicas, e não pela questão da representação científica *per se*. O discurso científico está repleto de passagens que parecem ser descrições de sistemas em uma disciplina específica, e as páginas de livros e periódicos estão repletas de discussões sobre as propriedades e o comportamento desses sistemas. Na mecânica, por exemplo, as propriedades dinâmicas de um sistema constituído por três esferas giratórias com distribuições homogêneas de massa são o foco de atenção, na biologia populações infinitas são investigadas, e na economia agentes perfeitamente racionais com acesso à informação perfeita trocam mercadorias. Apesar de sua estrutura superficial, ninguém confundiria as descrições de tais sistemas com descrições de um sistema **real**: sabemos muito bem que não existem tais sistemas.

Thomson-Jones (2010, p. 284) refere-se a essa descrição como uma "descrição de um sistema ausente". Essas descrições estão inseridas no que ele chama de "prática do valor de face" (2010, p. 285), a prática de falar e pensar sobre esses sistemas como se fossem reais. A prática do valor de face levanta uma série de questões. Que explicação deve ser dada a essas descrições e que tipo de objetos, se houver, eles descrevem? Estamos apresentando afirmações verificáveis ao apresentar descrições de sistemas ausentes?

A visão ficcional dos modelos fornece uma resposta: os modelos são semelhantes a lugares e personagens de ficção literária, e as afirmações sobre eles são verdadeiras ou falsas, da mesma forma que as afirmações sobre esses

lugares e personagens são verdadeiras ou falsas. Tal posição foi recentemente defendida explicitamente por alguns autores (FRIGG, 2010a, b; GODFREY-SMITH, 2006), mas não sem oposição (GIERE, 2009; MAGNANI, 2012). É importante notar que a analogia já existe há algum tempo (CARTWRIGHT, 1983; MCCLOSKEY, 1990; VAIHINGER, 1924). Isso deixa a espinhosa questão de como analisar personagens e lugares fictícios. Aqui, os filósofos da ciência podem recorrer a discussões da estética para preencher os detalhes sobre essas questões (FRIEND, 2007; SALIS, 2013).

A segunda versão da visão ficcional concentra-se explicitamente na representação. A maioria das teorias sobre a representação que encontramos até agora postula que existem sistemas-modelo e interpreta a representação científica como uma relação entre duas entidades, o sistema-modelo e o sistema-alvo. Toon chama isso de visão **indireta** da representação (2012, p.43). Na verdade, Weisberg vê esse caráter indireto como a característica definidora da modelagem (2007). Esta visão contrasta com o que Toon (2012, p. 43) e Levy (2015, p. 790) chamam de visão **direta** da representação. Essa visão não reconhece sistemas-modelo e visa, em vez disso, explicar a representação como uma forma de descrição direta. Nesta visão, os modelos fornecem uma "descrição imaginativa de coisas reais" (LEVY, 2012, p. 741), como o pêndulo real, e não há tal coisa como um sistema-modelo cuja descrição do pêndulo seja literalmente verdadeira (TOON, 2012, p. 43-44).

Toon articula a visão direta com base na teoria do faz de conta de Walton (1990). No cerne desta teoria está a noção de um jogo de faz de conta. Jogamos tal jogo se, por exemplo, ao caminharmos por uma floresta, imaginamos que os pedaços de tronco são ursos e, se avistamos um pedaço de tronco, imaginamos que avistamos um urso. Na terminologia de Walton, os pedaços de tronco são adereços, e a regra de que imaginamos um urso quando vemos um pedaço de tronco é um princípio de geração. Juntos, um adereço e um **princípio de geração** prescrevem o que deve ser imaginado. Walton considera uma grande variedade de adereços diferentes, incluindo estátuas e obras de ficção literária. Toon foca sobre um tipo particular de jogo em que somos aconselhados a imaginar algo de um objeto do mundo real. Uma estátua que mostra Napoleão a cavalo (TOON, 2012, p. 37) é um adereço que nos obriga a imaginar, por exemplo, que Napoleão tem uma certa fisionomia e certas expressões faciais. Ao ler *A Guerra dos Mundos* (2012, p. 39), devemos imaginar que a cúpula da Catedral de São Paulo foi atacada por alienígenas e, agora, tem um buraco no lado oeste.

O movimento crucial é dizer que os modelos são adereços em jogos de faz de conta. Especificamente, os modelos materiais são como a estátua de Napoleão e os modelos teóricos são como o texto de *A Guerra dos Mundos*: ambos aconselham, à sua maneira, imaginar algo sobre um objeto real. Um modelo bola e bastão de uma molécula de metano nos aconselha a imaginar coisas particulares sobre o metano, e um modelo descritivo que retrata um ponto de massa quicando em uma mola perfeitamente elástica representa a bola real e o sistema de mola, que prescreve imaginações sobre o sistema real. Isso fornece a seguinte resposta ao problema RE (TOON, 2012, p. 62):

> **Representação Direta**: **M** é uma representação científica de **T sse M** funciona como adereço no jogo do faz de conta que prescreve imaginações sobre **T**.

Essa explicação resolve alguns dos problemas apresentados na Seção 1: a **Representação Direta** é assimétrica, abre espaço para deturpações e, dadas suas raízes na estética, renuncia ao Problema da Demarcação. Essa visão absolve o Problema da Ontologia, uma vez que os modelos são objetos físicos ou descrições, e nenhum deles é problemático nesse contexto. Toon permanece calado sobre o Problema do Estilo e a aplicabilidade da matemática.

Questões importantes permanecem. De acordo com a **Representação Direta**, os modelos nos aconselham a imaginar certas coisas sobre seus sistemas-alvo. A explicação permanece silenciosa, no entanto, sobre a relação entre o que um modelo nos aconselha imaginar e o que um usuário do modelo deve realmente inferir sobre o sistema-alvo e, portanto, não oferece resposta para o problema RE. Levy (2015) identifica isso como uma lacuna na explicação de Toon e propõe preenchê-la, invocando a noção de "verdade parcial" de Yablo (2014), a ideia é que um usuário do modelo deve considerar as proposições imaginadas como parcialmente verdadeiras para seus sistemas-alvo. No entanto, como Levy admite, existem outros tipos de casos que não se encaixam no molde, principalmente as idealizações distorcidas. Isso requer um tratamento diferente e é uma questão em aberto saber qual tratamento seria esse.

Uma outra preocupação é como a **Representação Direta** lida com modelos sem alvo. Se não existe um sistema-alvo, então sobre o que o modelo aconselha imaginações? Toon conhece bem esses modelos e sugere a seguinte solução: se um modelo não tem um alvo, ele aconselha imaginações sobre um personagem

fictício (2012, p. 76). Essa solução, no entanto, possui custos ontológicos, e um dos objetivos declarados da visão direta era evitar esses custos, removendo os sistemas-modelo da imagem. Levy (2015) tem como objetivo recuperar a parcimônia ontológica e propõe um movimento radical: não há modelos sem alvo. Se um modelo (suposto) não tem alvo, então, ele não é um modelo. Resta uma dúvida, no entanto, de como essa visão pode ser enquadrada na prática científica, na qual modelos sem alvo não são apenas comuns, mas também claramente reconhecidos como tais.

7. Representação-Como

Na explicação de Goodman (1976) da representação estética, a ideia é que uma obra de arte não apenas denota seu sujeito, mas também o representa como sendo assim ou de outro modo. Elgin (2010) desenvolveu ainda mais essa explicação e, crucialmente, sugeriu que ela também se aplicasse às representações científicas. Primeiro, discutiremos a noção de representação-como de Goodman e Elgin e, em seguida, consideraremos uma extensão recente de sua estrutura.

7.1 Da Arte à Ciência

Muitos exemplos de representação epistêmica são exemplos do que Goodman e Elgin chamam de "representação-como". As caricaturas são exemplos paradigmáticos: Churchill é representado como um *bulldog* e Thatcher é representada como um boxeador. Mas a noção é mais geral: o *Retrato de Henrique VIII*, de Holbein, o representa como imponente e poderoso e a estátua de David Hume feita por Stoddart o representa como pensativo e sábio. Usando estas representações, podemos aprender sobre seus alvos, por exemplo, sobre a personalidade de um político ou de um filósofo. A ideia principal das visões discutidas nesta seção é que a representação científica funciona da mesma maneira. Um modelo do sistema solar o representa como contendo esferas perfeitas; o modelo logístico de crescimento representa a população como reproduzindo em intervalos fixos de tempo; e assim por diante. Em cada caso, os modelos podem ser usados para tentar aprender sobre seus alvos, determinando a forma como os primeiros representam a maneira de ser dos últimos.

A locução da representação-como funciona da seguinte maneira: um objeto **X** (por exemplo, uma imagem, estátua ou modelo) representa um sujeito **Y** (por

exemplo, uma pessoa ou sistema-alvo) como sendo assim (**Z**). A questão, então, é o que estabelece esse tipo de relacionamento representacional. A resposta requer a introdução de alguns dos conceitos que Goodman e Elgin usam para desenvolver sua explicação da representação-como.

Goodman e Elgin fazem uma distinção entre algo ser uma representação de um **Z** e algo ser uma representação-**Z** (Elgin 2010, p. 1-2; Goodman 1976, p. 21-26). Uma pintura de um unicórnio é uma representação de unicórnio porque mostra um unicórnio, mas não é uma representação de um unicórnio porque não há unicórnios. Ser uma representação-**Z** é um predicado de um lugar que categoriza as representações de acordo com seu tema. Ser uma representação *de* algo é estabelecido por denotação; é uma relação binária que se mantém entre um símbolo e o objeto que denota. Os dois podem, mas não precisam, coincidir. Algumas representações de cães são representações de cães, mas nem todas são (por exemplo, uma caricatura de Churchill), e nem todas as representações de cães são representações de cães (por exemplo, um raio pode representar o galgo mais rápido nas corridas).

A próxima noção é **exemplificação**: um objeto **X** exemplifica uma propriedade **P**, **sse X** instancia **P** e, assim, remete a **P** (GOODMAN, 1976, p. 53). No contexto atual, as propriedades devem ser entendidas no sentido mais amplo possível. Um item pode exemplificar propriedades de um local, propriedades de vários locais (ou seja, relações), propriedades de ordem superior, propriedades estruturais, etc. Exemplos paradigmáticos disso são amostras. Um fragmento de tinta no cartão de amostra de um fabricante instancia uma determinada cor e, ao mesmo tempo, se refere a essa cor (ELGIN, 1983, p. 71). Observe que a instanciação é necessária, mas insuficiente para a exemplificação: o cartão de amostra não exemplifica ser retangular, por exemplo. Quando um objeto exemplifica uma propriedade, ele nos fornece acesso epistêmico a essa propriedade.

A representação-como é, então, estabelecida pela combinação dessas noções: uma representação-**Z** exemplifica propriedades associadas a **Zs**[63] , e se a

[63] Muitas obras de arte não instanciam literalmente as propriedades que exemplificam. Retratos e estátuas não podem instanciar propriedades como velocidade e elegância; eles são feitos de papel ou bronze. Goodman e Elgin reconhecem isso e dizem que esses são exemplos de exemplificação metafórica (ELGIN, 1983, p. 81). Uma pintura pode, literalmente, instanciar a propriedade de ser cinza; pode, metaforicamente, instanciar a tristeza (GOODMAN, 1976, p. 50-52).

representação-**Z** denota **Y** adicionalmente, então, estas propriedades podem ser imputadas em **Y** (ELGIN, 2010, p. 10). Isso fornece a seguinte explicação da representação epistêmica:

> **Representação-Como**: **X** é uma representação epistêmica de **Y sse** (i) **X** denota **Y**, (ii) **X** é uma representação-**Z** que exemplifica as propriedades $\mathbf{P_1}$,..., $\mathbf{P_n}$ e (iii) **X** imputa $\mathbf{P_1}$,..., $\mathbf{P_n}$, ou propriedades relacionadas, em **Y**.

Aplicando isso no contexto científico, ou seja, permitindo que **X** alcance modelos e **Y** seus sistemas-alvo, chegamos a uma resposta para o problema RE. A **Representação-Como** também responde aos outros problemas apresentados na Seção 1: repudia o problema da demarcação e explica a direcionalidade da representação. É responsável pelo raciocínio substitutivo em termos das propriedades imputadas ao alvo. Se **Y** possui as propriedades imputadas, então, a representação é precisa, mas como o alvo não precisa necessariamente instanciá-las, isso permite a possibilidade de deturpação. Estilos diferentes podem ser explicados categorizando representações em termos de **Zs** diferentes ou em termos das propriedades que eles exemplificam. No entanto, pelo menos como declarado, a explicação permanece silenciosa sobre o problema da ontologia e a aplicabilidade da matemática. Discutiremos abaixo como explicar modelos sem alvos.

7.2 A Explicação DEKI

A **Representação-Como** levanta várias questões quando aplicada no contexto científico. O primeiro diz respeito à noção de representação-**Z**. Embora tenha apelo intuitivo no caso dos retratos, é menos claro como funciona no contexto da ciência. Phillips e Newlyn construíram um sistema elaborado de tubos e tanques para modelar uma economia (MORGAN; BOUMANS, 2004). Portanto, a máquina é uma representação de uma economia. Mas o que transforma um sistema de tubos e tanques em uma representação da economia?

Frigg e Nguyen (2016, p. 227-228) argumentam que, para transformar um objeto **X** em um modelo científico, ele deve ser interpretado da maneira apropriada (observe que eles não usam “interpretação” da maneira que Contessa o usa, como

discutido acima): propriedades que **X** possui, *qua* objeto, são emparelhadas com propriedades **Z** relevantes e propriedades quantitativas, isto é, propriedades como massa ou fluxo, que assumem valores numéricos, havendo a necessidade de uma associação adicional entre os valores da propriedade **X** e os valores da propriedade **Z** para a qual está mapeada. No caso da máquina Phillips-Newlyn, por exemplo, as propriedades hidráulicas da máquina estão associadas a propriedades econômicas e há uma regra que especifica que um litro de água corresponde a uma certa quantia da moeda do modelo econômico. As propriedades **X** e **Z** que estão tão associadas não precisam esgotar as propriedades que **X** instancia, nem todas as propriedades **Z** possíveis precisam estar associadas a uma propriedade **X**. Um modelo científico pode, então, ser definido como uma representação-**Z**, isto é, um objeto sob uma interpretação. Esta noção de modelo, explicitamente, não pressupõe um sistema-alvo e, portanto, abre espaço para modelos sem alvos. Os modelos podem, então, ser vistos como instanciando propriedades **Z sob a interpretação relevante**, o que explica como o modelo pode exemplificar tais propriedades.

A próxima questão é que as propriedades exemplificadas raramente são exatamente imputadas nos sistemas-alvo. De acordo com a **Representação-Como**, as propriedades imputadas são ou aquelas exemplificadas pela representação-**Z** "ou aquelas relacionadas". Frigg e Nguyen (2016, p. 228), com base em Frigg (2010a, p. 125-135), preferem ser explícitos sobre a relação entre as propriedades exemplificadas e aquelas que serão imputadas ao alvo. Eles fazem isso introduzindo uma “chave”, que associa explicitamente as propriedades exemplificadas às propriedades a serem imputadas ao alvo. Por exemplo, no caso de um mapa do metrô de Londres, a chave associa cores específicas a linhas de metrô específicas e, no caso de idealizações, a chave associa propriedades não idealizadas às propriedades do modelo.

A reunião das várias peças leva à seguinte explicação sobre a representação (FRIGG; NGUYEN, 2016, p. 229):

> DEKI: Seja **M** = <**X**, **I**> um modelo, onde **X** é um objeto e **I** uma interpretação. Seja **T** o sistema-alvo. **M** representa **T** como **Z** sse todas as condições seguintes forem satisfeitas:
>
> i. **M** denota **T**.
> ii. **M** exemplifica as propriedades **Z** $\mathbf{P}_1$, ..., $\mathbf{P}_n$.

iii. **M** vem com a chave **K** associando o conjunto {$\mathbf{P}_1$,…, $\mathbf{P}_n$} a um conjunto (possivelmente idêntico) de propriedades {$\mathbf{Q}_1$,…, $\mathbf{Q}_m$}.
iv. **M** imputa pelo menos uma das propriedades $\mathbf{Q}_1$, ..., $\mathbf{Q}_m$ a **T**.

M é uma representação científica de **T sse M** representa **T** como **Z** conforme definido em (i) – (iv).

O apelido "DEKI" destaca as principais características da explicação: denotação, exemplificação, chaveamento [*keying-up*] e imputação. DEKI responde aos problemas da Seção 1 da mesma maneira que a **Representação-Como** o faz. No entanto, faz acréscimos a esta última de, pelo menos, três maneiras. Primeiramente, as condições dadas deixam claro o que torna os modelos científicos representações-**Z** em primeiro lugar: as interpretações. Em segundo lugar, torna explícito que as propriedades exemplificadas pelo modelo não precisam ser imputadas exatamente ao alvo, e destaca a necessidade de investigar as chaves que especificam a relação entre as propriedades nos modelos e as propriedades que os modelos realmente imputam em seus alvos. Finalmente, torna explícito como explicar os modelos sem alvo. Um modelo científico que falha em denotar um sistema-alvo pode, no entanto, ser uma representação-**Z**. Um modelo de ponte que nunca foi construída ainda é uma representação de ponte que exemplifica propriedades relacionadas a pontes (estabilidade e assim por diante), apesar de não ser uma representação de nada. No entanto, como no caso da **Representação-Como**, as questões permanecem no que diz respeito ao problema da ontologia e à aplicabilidade da matemática.

Referências

ABELL, C. Canny Resemblance. **Philosophical Review**, v. 118, n. 2, p. 183-223, 2009. DOI:10.1215/00318108-2008-041.

AINSWORTH, P. Newman's Objection. **The British Journal for the Philosophy of Science**, v. 60, n. 1, p. 135-171, 2009. DOI:10.1093/bjps/axn051.

ANKENY, R. A.; LEONELLI, S. What's So Special About Model Organisms? **Studies in History and Philosophy of Science**, v. 42, n. 2, p. 313-323, 2011. DOI:10.1016/j.shpsa.2010.11.039.

ARONSON, J. L.; HARRÉ, R.; WAY, E. C. **Realism Rescued**: How Scientific Progress Is Possible. Chicago: Open Court, 1995.

BAILER-JONES, D. M. When Scientific Models Represent. **International Studies in the Philosophy of Science**, v. 17, n. 1, p. 59-74, 2003. DOI:10.1080/02698590305238.

BARTELS, A. Defending the Structural Concept of Representation. **Theoria**, v. 21, n. 1, p. 7-19, 2006.

BOGEN, J.; WOODWARD, J. Saving the Phenomena. **Philosophical Review**, v. 97, n. 3, p. 303-352, 1988. DOI:10.2307/2185445.

BOLINSKA, A. Epistemic Representation, Informativeness and the Aim of Faithful Representation. **Synthese**, v. 190, n. 2, p. 219-234, 2013. DOI:10.1007/s11229-012-0143-6.

BRADING, K.; LANDRY, E. Scientific Structuralism: Presentation and Representation. **Philosophy of Science**, v. 73, p. 571-581, 2006.

BRANDOM, R. B. **Making It Explicit**: Reasoning, Representing and Discursive Commitment. Cambridge: Harvard University Press, 1994.

BRANDOM, R. B. **Articulating Reasons**: An Introduction to Inferentialism, Cambridge: Harvard University Press, 2000.

BUDD, M. How Pictures Look. *In:* KNOWLES, D.; SKORUPSKI, J. (ed.). **Virtue and Taste.** Oxford: Blackwell, 1993, p. 154-175.

BUENO, O. Empirical Adequacy: A Partial Structure Approach. **Studies in the History and Philosophy of Science**, v. 28, n. 4, p. 585-610, 1997. DOI:10.1016/S0039-3681(97)00012-5.

BUENO, O. Models and Scientific Representations. *In:* MAGNUS, P.; BUSCH, J. (ed.). **New Waves in Philosophy of Science**. Basingstoke: Palgrave Macmillan, 2010. p. 94-111.

BUENO, O.; FRENCH, S. How Theories Represent. **The British Journal for the Philosophy of Science**, v. 62, n. 4, p. 857-894, 2011. DOI:10.1093/bjps/axr010.

BUENO, O.; FRENCH, S.; LADYMAN, J. On Representing the Relationship between the Mathematical and the Empirical. **Philosophy of Science**, v. 69, n. 3, p. 452-473, 2002. DOI:10.1086/342456.

CALLENDER, C.; COHEN, J. There Is No Special Problem About Scientific Representation. **Theoria**, v. 21, n. 1, p. 67-84, 2006.

CARTWRIGHT, N. **How the Laws of Physics Lie.** Oxford: Oxford University Press, 1983.

CARTWRIGHT, N. **The Dappled World**: A Study of the Boundaries of Science. Cambridge: Cambridge University Press, 1999.

CONTESSA, G. Scientific Representation, Interpretation, and Surrogative Reasoning. **Philosophy of Science**, v. 74, n. 1, p. 48-68, 2007. DOI:10.1086/519478.

CONTESSA, G. Scientific Models and Fictional Objects. **Synthese**, v. 172, p. 215-229, 2010. DOI:10.1007/s11229-009-9503-2.

CONTESSA, G. Scientific Models and Representation. *In:* FRENCH, S.; SAATSI, J. (ed.). **The Continuum Companion to the Philosophy of Science**. London: Continuum Press, 2011, p. 120-137.

DA COSTA, N. C. A.; FRENCH, S. The Model-Theoretic Approach to the Philosophy of Science. **Philosophy of Science**, v. 57, n. 2, p. 248-265, 1990. DOI:10.1086/289546.

DA COSTA, N. C. A.; FRENCH, S. **Science and Partial Truth**: A Unitary Approach to Models and Scientific Reasoning. Oxford: Oxford University Press, 2003.

DE DONATO RODRÍGUEZ, X.; ZAMORA BONILLA, J. Credibility, Idealisation, and Model Building: An Inferential Approach. **Erkenntnis**, 70, n. 1, p. 101-118, 2009. DOI:10.1007/s10670-008-9139-5v.

DECOCK, L.; DOUVEN, I. Similarity after Goodman. **Review of Philosophy and Psychology**, v. 2, n. 1, p. 61-75, 2011. DOI:10.1007/s13164-010-0035-y.

DONNELLAN, Keith S. Putting Humpty Dumpty Together Again. **Philosophical Review**, v. 77, n. 2, p. 203-215, 1968. DOI:10.2307/2183321.

DOWNES, Stephen M. Models, Pictures, and Unified Accounts of Representation: Lessons from Aesthetics for Philosophy of Science. **Perspectives on Science**, v. 17, p. 417-428, 2009.

DUCHEYNE, S. Scientific Representations as Limiting Cases. **Erkenntnis**, v. 76, n. 1, p. 73-89, 2012. DOI:10.1007/s10670-011-9309-8.

ELGIN, C. Z. **With Reference to Reference**. Indianapolis: Hackett, 1983.

ELGIN, C. Z. Telling Instances. *In:* FRIGG, R.; HUNTER, M. C. (ed.). **Beyond Mimesis and Convention**: Representation in Art and Science. New York: Springer, 2010, p. 1-18.

ELKINS, J. **The Domain of Images.** London: Cornell University Press, 1999.

FRENCH, S. A Model-Theoretic Account of Representation (or, I Don't Know Much About Art…But I Know It Involves Isomorphism). **Philosophy of Science**, v. 70, n. 5, p. 1472-1483, 2003. DOI:10.1086/377423.

FRENCH, S. **The Structure of the World. Metaphysics and Representation.** Oxford: Oxford University Press, 2014.

FRENCH, S.; LADYMAN, J. Reinflating the Semantic Approach. **International Studies in the Philosophy of Science**, v. 13, n. 2, p. 103-121, 1999. DOI:10.1080/02698599908573612.

FRIEND, S. Fictional Characters. **Philosophy Compass**, v. 2, n. 2, p. 141-156, 2007. DOI:10.1111/j.1747-9991.2007.00059.x.

FRIGG, R. Models and Representation: Why Structures Are Not Enough. *In:* Dietsch, P. (ed.). **Measurement in Physics and Economics Project Discussion Paper Series**. London: The London School of Economics and Political Science, 2002.

FRIGG, R. Scientific Representation and the Semantic View of Theories. **Theoria**, v. 21, n. 1, p. 49-65, 2006.

FRIGG, R. Fiction and Scientific Representation. 2010a. *In:* FRIGG, R.; HUNTER, M. C. (ed.). **Beyond Mimesis and Convention**: Representation in Art and Science. New York: Springer, 2010, p. 97-138.

FRIGG, R. Models and Fiction. **Synthese**, v. 172, p. 251-268, 2010b. DOI:10.1007/s11229-009-9505-0.

FRIGG, R.; HUNTER, M. C. (ed.). **Beyond Mimesis and Convention**: Representation in Art and Science. New York: Springer, 2010.

FRIGG, R.; NGUYEN, J. The Fiction View of Models Reloaded. **The Monist**, v. 99, n. 3, p. 225-242, 2016.

FRIGG, R.; NGUYEN, J.Models and Representation. In: MAGNANI, L.; BERTOLOTTI, T. (ed.). **Springer Handbook of Model-Based Science.** Springer: Cham, Switzerland, 2017. p. 49-102. DOI: 0.1007/978-3-319-30526-4_3.

GIERE, R. N. **Explaining Science**: A Cognitive Approach. Chicago: Chicago University Press, 1988.

GIERE, R. N. How Models Are Used to Represent Reality. **Philosophy of Science**, v. 71, n. 5, p. 742-752, 2004. DOI:10.1086/425063.

GIERE, R. N. Why Scientific Models Should Not Be Regarded as Works of Fiction. *In:* SUÁREZ, M. (ed.). Fictions in Science. **Philosophical Essays on Modelling and Idealization.** London: Routledge, 2009, p. 248-258.

GIERE, R. N. An Agent-Based Conception of Models and Scientific Representation. **Synthese**, v. 172, p. 269-281, 2010. DOI:10.1007/s11229-009-9506-z.

GLYMOUR, C. Theoretical Equivalence and the Semantic View of Theories. **Philosophy of Science**, v. 80, n. 2, p. 286-297, 2013. DOI:10.1086/670261.

GODFREY-SMITH, P. The Strategy of Model-Based Science. **Biology and Philosophy**, v. 21, p. 725-740, 2006.

GOODMAN, N. Seven Strictures on Similarity. *In:* GOODMAN, N. (ed.). **Problems and Projects.** New York: Bobs-Merril, 1972, p. 437-446.

GOODMAN, N. **Languages of Art**. 2. ed. Cambridge: Hackett, 1976.

HACKING, I. **Representing and Intervening**: Introductory Topics in the Philosophy of Natural Science. Cambridge: Cambridge University Press, 1983.

HALVORSON, H. What Scientific Theories Could Not Be. **Philosophy of Science**, v. 79, n. 2, p. 183-206, 2012. DOI:10.1086/664745.

HARTMANN, S. Models as a Tool for Theory Construction: Some Strategies of Preliminary Physics. *In:* HERFEL, W. E.; KRAJEWSKI, W.; NIINILUOTO, I.; WOJCICKI, R. (ed.). **Theories and Models in Scientific Processes**. Atlanta: Rodopi, 1995, p. 49-67. (Poznan Studies in the Philosophy of Science and the Humanities 44).

HODGES, W. **A Shorter Model Theory.** Cambridge: Cambridge University Press, 1997.

HUGHES, R. I. G. Models and Representation. **Philosophy of Science**, v. 64, p. S325-S36, 1997. DOI:10.1086/392611.

HUGHES, R. I. G. **The Theoretical Practises of Physics**: Philosophical Essays. Oxford: Oxford University Press, 2010.

KETLAND, J. Empirical Adequacy and Ramsification. **The British Journal for the Philosophy of Science**, v. 55, n. 2, p. 287-300, 2004. DOI:10.1093/bjps/55.2.287.

KNUUTTILA, T. Models, Representation, and Mediation. **Philosophy of Science**, v. 72, n. 5, p. 1260-1271, 2005. DOI:10.1086/508124.

KNUUTTILA, T. Modelling and Representing: An Artefactual Approach to Model-Based Representation. **Studies in History and Philosophy of Science**, v. 42, n. 2, p. 262-271, 2011.

KROES, P. Structural Analogies between Physical Systems. **The British Journal for the Philosophy of Science**, v. 40, n. 2, p. 145-154, 1989. DOI:10.1093/bjps/40.2.145.

LAURENCE, S.; MARGOLIS, E. Concepts and Cognitive Science. *In:* LAURENCE, S.; MARGOLIS, E. (ed.). **Concepts**: Core Readings. Cambridge: MIT Press, 1999, p. 3-81.

LEVY, A. Models, Fictions, and Realism: Two Packages. **Philosophy of Science**, v. 79, n. 5, p. 738-748, 2012. DOI:10.1086/667992.

LEVY, A. Modeling without Models. **Philosophical Studies**, v. 172, n. 3, p. 781-798, 2015. DOI:10.1007/s11098-014-0333-9.

LIU, C. Deflationism on Scientific Representation. *In:* Vassilios Karakostas and Dennis Dieks (eds.), EPSA Perspectives and Foundational Problems in Philosophy of Science, **Springer**, 93–102. 2013

LIU, C. Re-inflating the Conception of Scientific Representation. **International Studies in the Philosophy of Science**, v. 29, n. 1, p. 51-59, 2015. DOI:10.1080/02698595.2014.979671.

LLOYD, E. A Semantic Approach to the Structure of Population Genetics. **Philosophy of Science**, v. 51, n. 2, p. 242-264, 1984. DOI:10.1086/289179.

LOPES, D. **Understanding Pictures.** Oxford: Oxford University Press, 2004.

LYNCH, M.; WOOLGAR, S. **Representation in Scientific Practice.** Cambridge: MIT Press, 1990.

MACHOVER, M. **Set Theory, Logic and Their Limitations.** Cambridge: Cambridge University Press,1996.

MACKAY, A. F. Mr. Donnellan and Humpty Dumpty on Referring. **Philosophical Review**, v. 77, n. 2, p. 197-202, 1968. DOI:10.2307/2183320.

MAGNANI, L. Scientific Models Are Not Fictions: Model-Based Science as Epistemic Warfare. *In:* MAGNANI, L.; LI, P. (ed.). **Philosophy and Cognitive Science**: Western and Eastern Studies. Berlin-Heidelberg: Springer-Verlag, 2012, p. 1-38.

MAGNANI, L.; BERTOLOTTI, T. (ed.). **Springer Handbook of Model-Based Science**. New York: Springer, 2017.

MÄKI, U. Missing the World. Models as Isolations and Credible Surrogate Systems. **Erkenntnis**, v. 70, n. 1, p. 29-43, 2009. DOI:10.1007/s10670-008-9135-9.

MÄKI, U. Models and the Locus of Their Truth. **Synthese**, v. 180, n. 1, p. 47-63, 2011. DOI:10.1007/s11229-009-9566-0.

MCCLOSKEY, D. N. Storytelling in Economics. *In:* NASH, C. (ed.). **Narrative in Culture. The Uses of Storytelling in the Sciences Philosophy, and Literature.** London: Routledge, 1990, p. 5-22.

MORGAN, M.; BOUMANS, M. Secrets Hidden by Two-Dimensionality: The Economy as a Hydraulic Machine. *In:* Soraya de Chadarevian and Nick Hopwood (ed.). **Models**: The Third Dimension of Science, Stanford: Standford University Press, 2004, p. 369-401.

MORGAN, M.; MORRISON, M. **Models as Mediators**: Perspectives on Natural and Social Science. Cambridge: Cambridge University Press, 1999.

MORRISON, M. Models as Representational Structures. *In:* Stephan Hartmann, Carl Hoefer, and Luc Bovens (eds.). **Nancy Cartwright's Philosophy of Science.** New York: Routledge, 2008, p. 67-90.

MUNDY, B. **On the General Theory of Meaningful Representation**. Synthese, v. 67, n. 3, p. 391-437, 1986.

NEWMAN, M. H. A. Mr. Russell's Causal Theory of Perception. **Mind**, v. 37, n. 146, p. 137-148, 1928. DOI:10.1093/mind/XXXVII.146.137.

NGUYEN, J. On the Pragmatic Equivalence between Representing Data and Phenomena. **Philosophy of Science**, v. 83, n. 2, p. 171-191, 2016. DOI:10.1086/684959.

NIINILUOTO, I. Analogy and Similarity in Scientific Reasoning. *In:* HELMAN, D. H. (ed.). **Analogical Reasoning**: Perspectives of Artificial Intelligence. Cognitive Science, and Philosophy. Dordrecht: Kluwer, 1988, p. 271-298.

PARKER, W. S. Getting (Even More) Serious About Similarity. **Biology and Philosophy**, v. 30, n. 2, p. 267-276, 2015. DOI:10.1007/s10539-013-9406-y.

PERINI, L. Scientific Representation and the Semiotics of Pictures. *In:* MAGNUS, P. D.; BUSCH, J. (ed.). **New Waves in the Philosophy of Science.** New York: Macmilan, 2010, p. 131-154.

PERO, F.; SUÁREZ, M. Varieties of Misrepresentation and Homomorphism. **European Journal for Philosophy of Science**, v. 6, n. 1, p. 71-90, 2016. DOI:10.1007/s13194-015-0125-x.

PESCHARD, I. Making Sense of Modeling: Beyond Representation. **European Journal for Philosophy of Science**, v. 1, n. 3, p. 335-352, 2011. DOI:10.1007/s13194-011-0032-8.

PINCOCK, C. Overextending Partial Structures: Idealization and Abstraction. **Philosophy of Science**, v. 72, n. 5, p. 1248-1259, 2005. DOI:10.1086/508123.

PINCOCK, C. **Mathematics and Scientific Representation.** Oxford: Oxford University Press, 2012.

PORTIDES, D. Models and Theories. *In:* MAGNANI, L.; BERTOLOTTI, T. (ed.). **Springer Handbook of Model-Based Science**. New York: Springer, 2017, p. 25-48.

POZNIC, M. Representation and Similarity: Suárez on Necessary and Sufficient Conditions of Scientific Representation. **Journal for General Philosophy of Science**, v. 47, n. 2, p. 331-347, 2016. DOI:10.1007/s10838-015-9307-7.

PUTNAM, H. **Reason, Truth, and History**. Cambridge: Cambridge University Press, 1981.

PUTNAM, H. **The Collapse of the Fact-Value Distinction**. Cambridge: Harvard University Press, 2002.

QUINE, W.; ORMAN, O. **Ontological Relativity and Other Essays.** New York: Columbia University Press, 1969.

REDHEAD, M. The Intelligibility of the Universe. *In:* O'HEAR, A. (ed.). **Philosophy at the New Millennium.** Cambridge: Cambridge University Press, 2001, p. 73-90.

RESNIK, M. D. **Mathematics as a Science of Patterns.** Oxford: Oxford University Press, 1997.

RUSANEN, A-M.; LAPPI, O. An Information Semantic Account of Scientific Models. *In:* DE REGT, HENK W.; HARTMANN, S.; OKASHA, S. (ed.). **EPSA Philosophy of Science**. Amsterdam: Springer, 2012, p. 315-328.

SALIS, F. Fictional Entities. *In:* BRANQUINHO, J.; SANTOS, R. (ed.). **Online Companion to Problems in Analytic Philosophy**. Lisbon: Centre of Philosophy, University of Lisbon, 2013.

SHAPIRO, S. **Philosophy of Mathematics**: Structure and Ontology. Oxford: Oxford University Press, 1997.

SHAPIRO, S. **Thinking About Mathematics.** Oxford: Oxford University Press, 2000.

SHECH, E. Scientific Misrepresentation and Guides to Ontology: The Need for Representational Code and Contents. **Synthese**, v. 192, n. 11, p. 3463–3485, 2015. DOI:10.1007/s11229-014-0506-2.

SHEPARD, R. N. Multidimensional Scaling, Tree-Fitting, and Clustering. **Science**, v. 210, p. 390-398, 1980. DOI:10.1126/science.210.4468.390.

SUÁREZ, M. Scientific Representation: Against Similarity and Isomorphism. **International Studies in the Philosophy of Science**, v. 17, n. 3, p. 225-244, 2003. DOI:10.1080/0269859032000169442.

SUÁREZ, M. An Inferential Conception of Scientific Representation. **Philosophy of Science**, v. 71, n. 5, p. 767-779, 2004. DOI:10.1086/421415.

SUÁREZ, M. Deflationary Representation, Inference, and Practice. **Studies in History and Philosophy of Science**, v. 49, p. 36-47, 2015.

SUÁREZ, M.; SOLÉ, A. On the Analogy between Cognitive Representation and Truth. **Theoria**, v. 21, n. 1, p. 39-48, 2006.

SUPPES, P. A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences, **Suppes**, p. 10-23, 1960.

SUPPES, P. Models of Data. **Suppes**, p. 24–35, 1962.

SUPPES, P. **Studies in the Methodology and Foundations of Science**: Selected Papers from 1951 to 1969. Dordrecht: Reidel, 1969.

SWOYER, C. Structural Representation and Surrogative Reasoning. **Synthese**, v. 87, n. 3, p. 449-508, 1991. DOI:10.1007/BF00499820.

TEGMARK, M. The Mathematical Universe. **Foundations of Physics**, v. 38, n. 2, p. 101-150, 2008. DOI:10.1007/s10701-007-9186-9.

TELLER, P. Twilight of the Perfect Model Model. **Erkenntnis**, v. 55, n.3, p. 393-415, 2001. DOI:10.1023/A:1013349314515.

THOMSON-JONES, M. Missing Systems and Face Value Practice. **Synthese**, v. 172, n. 283-299, 2010. DOI:10.1007/s11229-009-9507-y.

THOMSON-JONES, M. Structuralism About Scientific Representation. *In:* BOKULICH, P. (ed.). **Scientific Structuralism**. Dordrecht: Springer, 2011, p. 119-141.

THOMSON-JONES, M. Modeling without Mathematics. **Philosophy of Science**, v. 79, n. 5, p. 761-772, 2012. DOI:10.1086/667876.

TOON, A. Models as Make-Believe. FRIGG, R.; HUNTER, M. C. (ed.). **Beyond Mimesis and Convention**: Representation in Art and Science. New York: Springer, 2010, p. 71-96.

TOON, A. Playing with Molecules. **Studies in History and Philosophy of Science**, v. 42, n. 4, p. 580-589, 2011.

TOON, A. **Models as Make-Believe. Imagination, Fiction and Scientific Representation**. Basingstoke: Palgrave Macmillan, 2012.

TVERSKY, A. Features of Similarity. **Psychological Review**, v. 84, n. 4, p. 327-352, 1977. DOI:10.1037/0033-295X.84.4.327.

UBBINK, J. B. Model, Description and Knowledge. **Synthese**, v. 12, n. 2, p. 302-319, 1960. DOI:10.1007/BF00485108.

VAIHINGER, H. **The Philosophy of as If**: A System of the Theoretical, Practical, and Religious Fictions of Mankind. London: Kegan Paul, 1924.

VAN FRAASSEN, B. C. **The Scientific Image**. Oxford: Oxford University Press, 1980.

VAN FRAASSEN, B. C. Structure and Perspective: Philosophical Perplexity and Paradox. *In:* CHIARA, M. L. D. (ed.). **Logic and Scientific Methods.** Dordrecht: Kluwer, 1997, p. 511-530.

VAN FRAASSEN, B. C. **The Empirical Stance.** London: Yale University Press, 2002.

VAN FRAASSEN, B. C. **Scientific Representation**: Paradoxes of Perspective, Oxford: Oxford University Press, 2008.

WALTON, Kendal L. **Mimesis as Make-Believe**: On the Foundations of the Representational Arts. Cambridge: Harvard University Press, 1990.

WEISBERG, M. Who Is a Modeler? **The British Journal for the Philosophy of Science**, v. 58, n. 2, p. 207-233, 2007. DOI:10.1093/bjps/axm011.

WEISBERG, M. Getting Serious about Similarity. **Philosophy of Science**, v. 59, n. 5, p. 785-794, 2012. DOI:10.1086/667845.

WEISBERG, M. **Simulation and Similarity**: Using Models to Understand the World. Oxford: Oxford University Press, 2013.

WIGNER, E. The Unreasonable Effectiveness of Mathematics in the Natural Sciences. **Communications on Pure and Applied Mathematics**, v. 13, n. 1-14, 1960.

YABLO, S. **Aboutness**. Princeton: Princeton University Press, 2014.

(VI) Modelos na ciência*

Autores: Roman Frigg e Stephan Hartmann
Tradução: Lucia C. Neco
Revisão: Renata Arruda

Modelos tem uma importância central em muitos contextos científicos. A centralidade de modelos como os modelos inflacionários em cosmologia, os modelos de circulação geral do clima global, o modelo de dupla hélice do DNA, os modelos evolutivos em biologia, os modelos baseados em agentes nas ciências sociais e os modelos de equilíbrio geral de mercados em seus respectivos domínios são exemplos disso. Os cientistas gastam uma quantidade significativa de tempo construindo, testando, comparando e revisando modelos. Além disso, grande parte das publicações científicas é dedicada à interpretação e discussão das implicações de modelos.

Como resultado, os modelos atraíram a atenção dos filósofos e agora existem literaturas de tamanhos consideráveis sobre vários aspectos da modelagem

---

* FRIGG, R; HARTMANN, S. Models in Science. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2020. Disponível em: https://plato.stanford.edu/archives/spr2020/entries/models-science/. Acesso em: 22 set. 2021.

The following is the translation of the entry on Models in Science by Roman Frigg and Stephan Hartmann, 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/spr2020/entries/models-science/. 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/models-science/. 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.

científica. Um resultado perceptível do envolvimento filosófico em modelos é uma proliferação de tipos de modelos reconhecidos na literatura filosófica. **Modelos de sondagem**, **modelos fenomenológicos**, **modelos computacionais**, **modelos de desenvolvimento**, **modelos explicativos**, **modelos empobrecidos**, **modelos de teste**, **modelos idealizados**, **modelos teóricos**, **modelos em escala**, **modelos heurísticos**, **modelos de caricatura**, **modelos exploratórios**, **modelos didáticos**, **modelos de fantasia**, **modelos mínimos**, **modelos de brinquedo**, **modelos imaginários**, **modelos matemáticos**, **modelos mecanicistas**, **modelos substitutos**, **modelos icônicos**, **modelos formais**, **modelos analógicos e modelos instrumentais** são apenas algumas das noções usadas para categorizar modelos. Embora à primeira vista essa abundância seja assustadora, ela pode ser controlada reconhecendo que essas noções se referem a diferentes problemas que surgem em conexão com os modelos. Modelos levantam perguntas sobre semântica (como os modelos representam? – se é que eles representam), ontologia (que tipo de coisas são modelos?), epistemologia (como aprendemos e explicamos com modelos?), e, é claro, em outros domínios dentro da filosofia da ciência.

1. Semântica: Modelos e Representação

Muitos modelos científicos são modelos representacionais: eles representam uma parte ou aspecto selecionado do mundo, que é o sistema-alvo do modelo. Exemplos muito conhecidos são o modelo atômico da bola de bilhar para gases, o modelo atômico de Bohr, o modelo de interação predador-presa de Lotka-Volterra, o modelo de Mundell-Fleming de uma economia aberta e o modelo em escala de uma ponte.

Isso levanta a questão sobre o que significa para um modelo representar um sistema-alvo. Esse problema é bastante complicado e se decompõe em vários subproblemas (para uma discussão aprofundada sobre a questão da representação, consulte o verbete sobre **representação científica**). Aqui, em vez de abordar a questão do que significa um modelo representar, focamos em vários tipos diferentes de representação que desempenham papéis importantes na prática da ciência baseada em modelos, a saber, modelos em escala, modelos analógicos, modelos idealizados, modelos de brinquedo, modelos mínimos, modelos fenomenológicos, modelos exploratórios e modelos de dados. Essas categorias não são mutuamente exclusivas e um determinado modelo pode se enquadrar em várias categorias ao mesmo tempo.

Modelos em escala. Alguns modelos são cópias em tamanho reduzido ou ampliadas de seus sistemas-alvo (BLACK, 1962). Um exemplo típico é um pequeno carro de madeira que é colocado em um túnel de vento para explorar as propriedades aerodinâmicas do carro. A intuição é que um modelo em escala é uma réplica naturalista ou uma imagem espelhada verdadeira do alvo; por esse motivo, os modelos em escala também são chamados de "modelos verdadeiros" (ACHINSTEIN, 1968, cap. 7). No entanto, não existe um modelo em escala perfeitamente fiel; a fidelidade é sempre restrita a alguns aspectos. A maquete em madeira do carro fornece uma representação fiel da forma do carro, mas não do seu material. E mesmo nos aspectos em que um modelo é uma representação fiel, a relação entre propriedades do modelo e propriedades do alvo geralmente não é direta. Quando os engenheiros usam, digamos, um modelo em escala de 1:100 de um navio para investigar a resistência que um navio real experimenta ao se mover pela água, eles não podem simplesmente medir a resistência que o modelo experimenta e depois multiplicá-lo pela escala. De fato, a resistência enfrentada pelo modelo não se traduz na resistência enfrentada pelo navio real de maneira direta (ou seja, não se pode simplesmente estimar a resistência à água apenas com a escala do modelo: a resistência do navio real não é necessariamente cem vezes a resistência à água do modelo 1:100). As duas quantidades mantêm uma relação não-linear complicada entre si, e a fórmula exata dessa relação é geralmente não trivial e surge como resultado de um estudo aprofundado da situação (STERRETT, 2006, no prelo; PINCOCK, no prelo).

Modelos analógicos. Exemplos padrão de modelos analógicos incluem o modelo da bola de bilhar de um gás, o modelo hidráulico de um sistema econômico e o modelo de um buraco negro "sônico" [*dumb hole*]. No nível mais básico, duas coisas são análogas se houver certas semelhanças relevantes entre elas. Em um texto clássico, Hesse (1963) distingue diferentes tipos de analogias de acordo com os tipos de relações de semelhança estabelecidas entre dois objetos. Um tipo simples de analogia é aquela baseada em propriedades compartilhadas. Existe uma analogia entre a Terra e a Lua, baseada no fato de que ambos são corpos esféricos, sólidos, opacos e grandes que recebem calor e luz do sol, giram em torno de seus eixos e gravitam em direção a outros corpos. Mas a equidade das propriedades não é uma condição necessária. Uma analogia entre dois objetos também pode ser baseada em semelhanças relevantes entre suas propriedades. Nesse sentido mais liberal, podemos dizer que existe uma analogia entre som e luz, porque os ecos são semelhantes às reflexões, a sonoridade ao brilho, a altura do tom à cor,

a detectabilidade pelo ouvido à detectabilidade pelo olho, e assim por diante.

As analogias também podem ser baseadas na similaridade ou semelhança das relações entre partes de dois sistemas, e não nas suas propriedades monádicas. É nesse sentido que se afirma que a relação de um pai com seus filhos é análoga à relação do estado com seus cidadãos. As analogias mencionadas até agora são o que Hesse chama de "analogias materiais". Obtemos uma noção mais formal de analogia quando abstraímos das características concretas dos sistemas e focamos apenas em sua configuração formal. O que o modelo analógico compartilha com seu sistema-alvo não é um conjunto de características, mas o mesmo padrão de relações abstratas (ou seja, a mesma estrutura, onde a estrutura é entendida em sentido formal). Essa noção de analogia está intimamente relacionada ao que Hesse chama de "analogia formal". Dois itens são relacionados por analogia formal se ambos são interpretações do mesmo cálculo formal. Por exemplo, existe uma analogia formal entre um pêndulo oscilante e um circuito elétrico oscilante porque ambos são descritos pela mesma equação matemática.

Outra distinção importante devido a Hesse é aquela entre analogias positivas, negativas e neutras. A analogia positiva entre dois itens consiste nas propriedades ou relações que eles compartilham (moléculas de gás e bolas de bilhar têm massa); a analogia negativa consiste nas propriedades que eles não compartilham (bolas de bilhar são coloridas, moléculas de gás não); a analogia neutra compreende as propriedades das quais (ainda) não se sabe se pertencem à analogia positiva ou negativa (as bolas de bilhar e moléculas têm a mesma seção transversal nos processos de dispersão?). As analogias neutras desempenham um papel importante na pesquisa científica porque dão origem a perguntas e sugerem novas hipóteses. Por esse motivo, vários autores enfatizaram o papel heurístico que as analogias desempenham na teoria e na construção de modelos, bem como no pensamento criativo (BAILER-JONES; BAILER-JONES, 2002; BAILER-JONES, 2009, cap. 3; HESSE, 1974; HOLYOAK; THAGARD, 1995; KROES, 1989; PSILLOS, 1995; HELMAN, 1988).

Também foi discutido se o uso de modelos analógicos pode, em alguns casos, ser confirmatório no sentido bayesiano. Hesse (1974, p. 208-219) argumenta que isso é possível se a analogia for uma analogia material. Bartha (2010, 2019) discorda e argumenta que os modelos analógicos não podem ser confirmatórios no sentido bayesiano porque as informações encapsuladas em um modelo analógico fazem parte do conhecimento de fundo relevante, o que tem como consequência a probabilidade posterior de uma hipótese sobre um sistema-alvo não pode mudar

como resultado da observação da analogia. Os modelos analógicos podem, portanto, apenas estabelecer a plausibilidade de uma conclusão no sentido de justificar uma atribuição de probabilidade anterior não desprezível (BARTHA, 2010, §8.5).

Mais recentemente, essas questões foram discutidas no contexto das chamadas experiências analógicas, que prometem fornecer conhecimento sobre um sistema-alvo experimentalmente inacessível (por exemplo, um buraco negro) manipulando outro sistema, o sistema de teste (por exemplo, um Condensado de Bose-Einstein). Dardashti, Thébault e Winsberg (2017) e Dardashti, Hartmann e colaboradores (2019) argumentaram que, dadas certas condições, uma simulação analógica de um sistema por outro sistema pode confirmar reivindicações sobre o sistema-alvo (por exemplo, que buracos negros emitem radiação Hawking). Veja Crowther *et al*. (a ser publicado) para uma discussão crítica e também o verbete sobre simulações computacionais na ciência.

Modelos idealizados. Modelos idealizados são modelos que envolvem uma simplificação deliberada ou distorção de algo complexo com o objetivo de torná-lo mais simples ou compreensível. Planos sem atrito, massas pontuais, sistemas completamente isolados, agentes oniscientes e totalmente racionais e mercados em perfeito equilíbrio são exemplos bem conhecidos. As idealizações são um meio crucial para a ciência lidar com sistemas que são muito difíceis de estudar em toda a sua complexidade (POTOCHNIK, 2017).

Os debates filosóficos sobre a idealização concentraram-se em dois tipos gerais de idealizações: as chamadas idealizações aristotélicas e galileanas. A idealização aristotélica equivale a "arrancar", em nossa imaginação, todas as propriedades de um objeto concreto que acreditamos não serem relevantes para o problema em questão. Há divergências sobre como isso é feito. Jones (2005) e Godfrey-Smith (2009) oferecem uma análise da abstração em termos de verdade: enquanto uma abstração se omite sobre certas características ou aspectos do sistema, ela não diz nada falso e ainda oferece uma descrição verdadeira (embora restrita). Isso permite que os cientistas se concentrem em um conjunto limitado de propriedades isoladas. Um exemplo é um modelo de mecânica clássica do sistema planetário, que descreve a posição de um objeto em função do tempo e desconsidera todas as outras propriedades dos planetas. Cartwright (1989, cap. 5), Musgrave (1981), que usa o termo "suposições de insignificância", e Mäki (1994), que fala do "método de isolamento", permitem que as abstrações digam algo falso, por exemplo, omitindo um fator causalmente relevante.

As idealizações galileanas são aquelas que envolvem distorções deliberadas: os físicos constroem modelos que consistem em massas pontuais que se deslocam em planos sem atrito; os economistas presumem que os agentes são oniscientes; biólogos estudam populações isoladas; e assim por diante. Usar simplificações desse tipo sempre que uma situação é muito difícil de resolver era característico da abordagem de Galileu à ciência. Por esta razão, é comum se referir a idealizações "distorcidas" desse tipo como "idealizações Galileanas" (MCMULLIN, 1985). Um exemplo para tal idealização é um modelo de movimento em uma pista de gelo que assume que o gelo é desprovido de atrito, quando, na realidade, ele tem atrito baixo, mas não nulo.

As idealizações galileanas são algumas vezes caracterizadas como idealizações controladas, isto é, aquelas que permitem a des-idealização por remoção sucessiva das suposições distorcidas (MCMULLIN, 1985; WEISBERG, 2007). Assim interpretadas, as idealizações galileanas não cobrem todas as idealizações distorcidas. Batterman (2002, 2011) e Rice (2015, 2019) discutem idealizações distorcidas que são ineliminâveis, pois não podem ser removidas do modelo sem desmontá-lo por completo.

O que um modelo envolvendo distorções nos diz sobre a realidade? Laymon (1991) formulou uma teoria que entende as idealizações como limites ideais: imagine uma série de refinamentos da situação real que se aproximam do limite postulado, e então exija que quanto mais próximas as propriedades de um sistema chegam do limite ideal, mais próximo seu comportamento chegará ao comportamento do sistema no limite (monotonicidade). Se for esse o caso, os cientistas podem estudar o sistema no limite e transportar as conclusões desse sistema para sistemas distantes do limite. Mas essas condições nem sempre valem. Na verdade, pode acontecer que o sistema limitador não se aproxime do sistema no limite. Se isso acontecer, estaremos diante de um limite singular (BERRY, 2002). Nesses casos, o sistema no limite pode exibir um comportamento diferente do comportamento dos sistemas distantes do limite. Limites desse tipo aparecem em vários contextos, mais notadamente na teoria das transições de fase na mecânica estatística. No entanto, não há consenso sobre a interpretação correta de tais limites. Batterman (2002, 2011) os vê como indicativos de fenômenos emergentes, enquanto Butterfield (2011a,b) os vê como compatíveis com a redução.

As idealizações galileanas e aristotélicas não são mutuamente exclusivas, e muitos modelos exibem as duas, ou seja, eles levam em conta um conjunto restrito de propriedades e as distorcem. Considere novamente o modelo da mecânica clássica do sistema planetário: o modelo leva em consideração apenas um conjunto

restrito de propriedades e as distorce, por exemplo, descrevendo planetas como esferas ideais com uma distribuição de massa simétrica à rotação.

Um conceito intimamente relacionado à idealização é a aproximação. Em um sentido amplo, **A** pode ser chamado de aproximação de **B** se **A** estiver de alguma forma perto de **B**. Isso, entretanto, é muito abrangente, porque abre espaço para qualquer semelhança ser qualificada como uma aproximação. Rueger e Sharp (1998) limitam as aproximações à proximidade **quantitativa** e Portides (2007) enquadra-a como um conceito essencialmente matemático. Nessa noção, **A** é uma aproximação de **B**, se e somente se, **A** estiver próximo de **B** em um sentido matemático especificável, em que o sentido relevante de "próximo" será dado pelo contexto. Um exemplo é a aproximação de uma curva com outra, que pode ser conseguida expandindo uma função em uma série de potências e mantendo apenas os primeiros dois ou três termos. Em diferentes situações, aproximamos uma equação de outra deixando um parâmetro de controle tender para zero (REDHEAD, 1980). Isso levanta a questão de como as aproximações são diferentes das idealizações, que também podem envolver proximidade matemática. Norton (2012) entende a distinção entre os dois como referencial: uma aproximação é uma descrição inexata do alvo, enquanto uma idealização introduz um sistema secundário (real ou fictício) que representa o sistema-alvo (sendo distinto dele). Se dissermos que o período do pêndulo na parede é de aproximadamente dois segundos, então isso é uma aproximação; se raciocinarmos sobre o pêndulo real assumindo que o peso (bob) do pêndulo é uma massa pontual e que a corda não tem massa (ou seja, se assumirmos que o pêndulo é o chamado pêndulo ideal), então usamos uma idealização. Separar idealizações e aproximações dessa maneira não implica que não possa haver relações interessantes entre as duas. Por exemplo, uma aproximação pode ser justificada dizendo que ela é a expressão matemática de uma idealização aceitável (por exemplo, quando omitimos um termo dissipativo em uma equação de movimento porque assumimos a idealização de que o sistema não tem atrito).

Modelos de brinquedo. Os modelos de brinquedo são representações extremamente simplificadas e fortemente distorcidas de seus alvos, e frequentemente representam apenas um pequeno número de fatores causais ou explicativos (HARTMANN, 1995; REUTLINGER *et al*. 2018; NGUYEN, no prelo). Exemplos típicos são o modelo Lotka–Volterra em ecologia populacional (WEISBERG, 2013) e o modelo de segregação de Schelling nas ciências sociais (SUGDEN, 2000). Os modelos de brinquedo geralmente não têm um bom desempenho em termos de predição e adequação empírica e parecem servir a outros objetivos epistêmicos

(mais sobre isso na Seção 3). Isso levanta uma questão sobre se eles devem ser realmente considerados representacionais (LUCZAK, 2017).

Alguns modelos de brinquedo são caracterizados como "caricaturas" (GIBBARD; VARIAN, 1978; BATTERMAN; RICE, 2014). Modelos de caricatura isolam um pequeno número de características salientes de um sistema e as distorcem em um caso extremo. Um exemplo clássico é o modelo de mercado de automóveis ("o mercado de limões") de Akerlof (1970), que explica a diferença de preço entre carros novos e usados apenas em termos de informação assimétrica, desconsiderando todos os outros fatores que podem influenciar os preços de carros (SUGDEN, 2000). No entanto, é controverso se esses modelos altamente idealizados ainda podem ser considerados representações informativas de seus sistemas-alvo. Para uma discussão de modelos de caricatura, particularmente em economia, ver Reiss (2006).

Modelos mínimos. Os modelos mínimos estão intimamente relacionados aos modelos de brinquedo, pois também são altamente simplificados. Eles são tão simplificados que alguns argumentam que não são representacionais: eles não têm qualquer similaridade, isomorfismo ou relação de semelhança com o mundo (BATTERMAN; RICE, 2014). Argumenta-se que muitos modelos econômicos são desse tipo (GRÜNE-YANOFF, 2009). Os modelos econômicos mínimos também não são restritos pelas leis naturais e não isolam nenhum fator real (*ibid*.). Porém, modelos mínimos nos ajudam a aprender algo sobre o mundo no sentido de que funcionam como substitutos de um sistema real: os cientistas podem estudar o modelo para aprender algo sobre o seu alvo. No entanto, não é consenso se modelos mínimos podem ajudar os cientistas a aprender algo sobre o mundo se eles não representam nada (FUMAGALLI, 2016). Modelos mínimos que supostamente carecem de qualquer semelhança ou representação também são usados em diferentes partes da física para explicar o comportamento em macroescala de vários sistemas cujo comportamento em microescala é extremamente diverso (BATTERMAN; RICE, 2014; RICE, 2018, 2019; SHECH, 2018). Exemplos típicos são os atributos de transições de fase e o fluxo de fluidos. Os defensores dos modelos mínimos argumentam que o que fornece uma explicação do comportamento em macroescala de um sistema nesses casos não é uma característica que o sistema e o modelo têm em comum, mas o fato de que o sistema e o modelo pertencem à mesma **classe de universalidade** (uma classe de modelos que exibem o mesmo comportamento limitante, embora mostrem comportamentos muito diferentes em escalas finitas). É, no entanto, discutível se explicações desse tipo são possíveis sem referência a pelo menos algumas características comuns (LANGE, 2015; REUTLINGER, 2017).

Modelos fenomenológicos. Modelos fenomenológicos foram definidos de maneiras diferentes, embora relacionadas. Uma definição comum considera que estes são modelos que representam apenas propriedades observáveis de seus alvos e se abstêm de postular sobre mecanismos ocultos ou outros semelhantes (BOKULICH, 2011). Outra abordagem, por McMullin (1968), define modelos fenomenológicos como modelos que são independentes de teorias. Isso, no entanto, parece ser muito forte. Muitos modelos fenomenológicos, embora não possam ser derivados de uma teoria, incorporam princípios e leis associadas a teorias. O modelo da gota líquida do núcleo atômico, por exemplo, retrata o núcleo como uma gota líquida e o descreve com diversas propriedades (tensão superficial e carga, entre outras) originadas em diferentes teorias (hidrodinâmica e eletrodinâmica, respectivamente). Certos aspectos dessas teorias – embora geralmente não as teorias completas – são usados para determinar as propriedades estáticas e dinâmicas do núcleo. Finalmente, é tentador identificar modelos fenomenológicos como **modelos de um fenômeno**. Aqui, "fenômeno" é um termo genérico que cobre todas as características relativamente estáveis e gerais do mundo que são interessantes do ponto de vista científico. O enfraquecimento do som em função da distância até a fonte, o decaimento das partículas alfa, as reações químicas que ocorrem quando um pedaço de calcário se dissolve em um ácido, o crescimento de uma população de coelhos e a dependência dos preços de casas sobre a taxa básica da Reserva Federal [*Federal Reserve*] são fenômenos nesse sentido. Para uma discussão mais aprofundada, consulte Bailer-Jones (2009, cap. 7), Bogen e Woodward (1988).

Modelos exploratórios. Modelos exploratórios são modelos que não são propostos em primeiro lugar para aprender algo sobre um sistema-alvo específico ou um fenômeno particular estabelecido experimentalmente. Os modelos exploratórios funcionam como ponto de partida para futuras explorações nas quais o modelo é modificado e refinado. Gelfert (2016) aponta que modelos exploratórios podem fornecer provas de conceitos [*proofs-of-principle* ou *proofs-of-concept*] e sugerir explicações do que possivelmente aconteceu [*how-possibly explanations*] (2016, cap. 4). Como exemplo, Gelfert menciona os primeiros modelos em ecologia teórica, como o modelo Lotka-Volterra de interação predador-presa, que imita o comportamento qualitativo de aceleração e desaceleração no crescimento populacional em um ambiente com recursos limitados (2016, p.80). Esses modelos não fornecem uma descrição precisa do comportamento de qualquer população real, mas fornecem o ponto de partida para o desenvolvimento de modelos mais realistas. Massimi (2019)

observa que os modelos exploratórios fornecem conhecimento modal. Fisher (2006) vê esses modelos como ferramentas para o exame das características de uma dada teoria.

Modelos de dados. Um modelo de dados é uma versão corrigida, retificada, organizada e, em muitos casos, idealizada dos dados que obtemos com a observação imediata, os chamados dados brutos (SUPPES, 1962). Normalmente, elimina-se primeiro os erros (por exemplo, removem-se pontos do registro que são resultado de observações incorretas) e, em seguida, apresenta-se os dados de uma forma "limpa", por exemplo, desenhando uma curva suave através de um conjunto de pontos. Essas duas etapas são comumente chamadas de "redução de dados" e "ajuste de curvas". Quando investigamos, por exemplo, a trajetória de um determinado planeta, primeiro eliminamos pontos enganosos dos registros de observação e depois ajustamos uma curva suave aos restantes. Os modelos de dados desempenham um papel crucial na confirmação de teorias, porque é em relação ao **modelo** de dados, e não aos dados brutos, geralmente confusos e complexos, que as teorias são testadas.

A construção de um modelo de dados pode ser extremamente complicada. Ela requer técnicas estatísticas sofisticadas e levanta sérias questões metodológicas e filosóficas. Como decidimos quais pontos do registro precisam ser removidos? E, dado um conjunto limpo de dados, que curva ajustamos a ele? A primeira questão foi tratada principalmente no contexto da filosofia do experimento (GALISON, 1997; STALEY, 2004). No cerne da última questão está o chamado problema de ajuste de curvas, que é o de que os dados brutos não ditam nem a forma da curva ajustada nem quais técnicas estatísticas os cientistas deveriam usar para construir esta curva. A escolha e a racionalização das técnicas estatísticas são o assunto da filosofia da estatística, e a Bandyopadhyay e Forster (2011) para uma discussão sobre essas questões. Discussões adicionais de modelos de dados podem ser encontradas em Bailer-Jones (2009, cap. 7), Brewer e Chinn (1994), Harris (2003), Hartmann (1995), Laymon (1982), Mayo (1996, 2018), e Suppes (2007).

A coleta, o processamento, a disseminação, a análise, a interpretação e o armazenamento de dados levantam muitas questões importantes além das questões estritamente relativas aos modelos de dados. Leonelli (2016, 2019) investiga o status dos dados na ciência, argumenta que os dados devem ser definidos não por sua procedência, mas por sua função como evidência e estuda como os dados viajam entre diferentes contextos.

2. Ontologia: O que são modelos?

O que são modelos? Ou seja, com que tipo de objeto os cientistas estão lidando quando trabalham com um modelo? Vários autores são céticos de que essa questão tenha alguma resposta significativa, porque os modelos não pertencem a uma categoria ontológica distinta e qualquer coisa pode ser um modelo (CALLENDER; COHEN, 2006; GIERE, 2010; SUÁREZ, 2004; SWOYER, 1991; TELLER, 2001). Contessa (2010) responde que isso é um *non sequitur*. Mesmo que, de um ponto de vista ontológico, qualquer coisa possa ser um modelo e a classe de coisas que são referidas como modelos contenha uma coleção heterogênea de coisas diferentes, isso não significa que seja impossível ou inútil desenvolver uma ontologia de modelos. Isso ocorre porque, mesmo que nem todos os modelos sejam de um tipo ontológico particular, pode-se, no entanto, perguntar a quais tipos ontológicos pertencem as coisas que *de facto* são usadas como modelos. Pode haver vários desses tipos e cada tipo pode ser analisado em seu próprio contexto. Os tipos de objetos que os cientistas usam como modelos têm repercussões importantes sobre como os modelos desempenham funções relevantes, como representação ou explicação, e, portanto, essa questão não pode ser descartada como "apenas sociologia".

Os objetos que comumente servem como modelos pertencem, de fato, a diferentes tipos ontológicos: objetos físicos, objetos fictícios, objetos abstratos, estruturas de teoria dos conjuntos, descrições, equações ou combinações de alguns destes, são frequentemente chamados de modelos, e alguns modelos podem cair em ainda outras classes de coisas. Seguindo o conselho dado por Contessa, o objetivo é desenvolver uma ontologia para cada um deles. Aqueles com interesse em ontologia podem ver isso como um objetivo em si. Vale a pena observar, no entanto, que a questão tem repercussões além da ontologia e depende da forma como se entende a semântica e a epistemologia dos modelos.

2.1 Objetos físicos

Alguns modelos são objetos físicos. Esses modelos são comumente chamados de "modelos físicos". Exemplos padrão de modelos desse tipo são modelos em escala de objetos como pontes e navios (*vide* Seção 1), o modelo de DNA de Watson e Crick (SCHAFFNER, 1969), o modelo hidráulico de economia de Phillips e Newlyn (MORGAN; BOUMANS, 2004), o Modelo da Baía de São Francisco

feita pelo Corpo de Engenheiros do Exército dos EUA (WEISBERG, 2013), o modelo da mioglobina feito de plasticina de Kendrew (FRIGG; NGUYEN, 2016) e organismos-modelo nas ciências da vida (LEONELLI; ANKENY, 2012; LEONELLI, 2010; LEVY; CURRIE, 2015). Todos esses são objetos físicos que servem de modelos. Os modelos físicos não dão origem a novas dificuldades ontológicas além dos problemas bem conhecidos de conexão com objetos já discutidos pelos metafísicos. Por exemplo, os problemas que dizem respeito à natureza das propriedades, a identidade de objetos, a relação parte e todo, e assim por diante.

No entanto, muitos modelos não são modelos físicos. O modelo do átomo de Bohr, um pêndulo sem fricção ou uma população isolada, por exemplo, estão na mente do cientista, e não no laboratório e não precisam ser fisicamente realizados e experimentados para servir de modelos. Esses modelos "não físicos" levantam sérias questões ontológicas, e não há consenso de como eles deveriam ser analisados da melhor forma. No restante desta seção, revisamos algumas das sugestões que recentemente chamaram a atenção na literatura sobre modelos.

2.2 Objetos fictícios e objetos abstratos

O que se tornou conhecido como a **visão ficcional dos modelos** vê os modelos como semelhantes aos objetos imaginários da ficção literária - isto é, semelhantes a personagens fictícios como Sherlock Holmes ou lugares fictícios como a Terra Média (GODFREY-SMITH, 2007). Portanto, quando Bohr apresentou seu modelo do átomo, ele introduziu um objeto fictício do mesmo tipo que o objeto que Conan Doyle apresentou quando inventou Sherlock Holmes. Essa visão se encaixa bem com a prática científica, onde os cientistas muitas vezes falam sobre modelos como se fossem objetos e muitas vezes se consideram estar descrevendo átomos, populações ou economias **imaginárias**. Ela também se encaixa bem com visões filosóficas que entendem a construção e manipulação de modelos como aspectos essenciais da investigação científica (MORGAN, 1999), mesmo que os modelos não sejam objetos físicos, porque essas práticas parecem estar direcionadas a **algum** tipo de objeto.

Quais questões filosóficas são resolvidas por esse movimento? O discurso ficcional e as entidades ficcionais enfrentam questões filosóficas bem conhecidas e pode-se argumentar que simplesmente comparar modelos a ficções, equivale a explicar *obscurum per obscurius*. Uma maneira de contrariar essa objeção e motivar

a visão ficcional dos modelos é apontar o poder heurístico dessa visão. Nesse sentido, Frigg (2010b) identifica cinco questões específicas que uma ontologia de modelos deve abordar e, em seguida, observa que essas questões surgem de maneiras muito semelhantes na discussão sobre a ficção (as questões são as condições de identidade, a atribuição de propriedade, a semântica de declarações, as condições de verdade e a epistemologia de objetos imaginários). Comparar os modelos à ficção tem valor heurístico porque existe uma rica literatura sobre ficção que oferece várias soluções para essas questões.

Apenas uma pequena parte das opções disponíveis na extensa literatura sobre ficção foi realmente explorada no contexto de modelos científicos. Contessa (2010) formula o que chama de "explicação dualista", segundo a qual um modelo é um objeto abstrato que representa um possível objeto concreto. O modelo do átomo de Rutherford, por exemplo, é um objeto abstrato que atua como substituto para um dos possíveis sistemas que contêm um elétron orbitando em torno de um núcleo em uma órbita bem definida. Barberousse e Ludwig (2009) e Frigg (2010b) seguem um caminho diferente e desenvolvem uma descrição dos modelos como ficções com base na teoria do faz de conta [*pretense theory of fiction*] de Walton (1990). De acordo com essa visão, as sentenças de uma passagem de texto que introduz um modelo devem ser vistas como um suporte em um jogo de faz de conta, e o modelo é o produto de um ato de imaginação [*pretense*]. Esta é uma posição antirrealista na medida em que considera falar de "objetos" modelo como figuras de linguagem porque, em última análise, não existem objetos-modelo, os modelos vivem apenas na imaginação dos cientistas. Salis (no prelo) reformula essa visão para se tornar o que ela chama de "a nova visão ficcional dos modelos". A principal diferença reside no fato de que o que é considerado o modelo são as descrições do modelo e seu conteúdo, e não as fantasias que eles determinam. Esta é uma visão realista dos modelos, porque as descrições existem.

A visão ficcional não é isenta de críticas. Giere (2009), Magnani (2012), Pincock (2012), Portides (2014) e Teller (2009) rejeitam a abordagem ficcional e argumentam, de diferentes maneiras, que os modelos não devem ser considerados ficções. Weisberg (2013) defende uma posição intermediária que entende as ficções como desempenhando um papel heurístico, mas nega que devam ser consideradas como parte de um modelo científico. O núcleo comum dessas críticas é que a visão ficcional interpreta mal a posição epistêmica dos modelos. Chamar algo de ficção, de acordo com as críticas, é o mesmo que dizer que é falso, e é injustificado chamar um modelo inteiro de falso – e, portanto, afirmar que ele falha em capturar como o

mundo é – só porque o modelo envolve certas suposições falsas ou elementos fictícios. Em outras palavras, uma representação não é automaticamente considerada ficção apenas porque possui algumas imprecisões. Os defensores da visão ficcional concordam com este ponto, mas negam que a noção de ficção deva ser analisada em termos de falsidade. O que torna uma obra uma ficção não é sua falsidade (ou alguma proporção de afirmações falsas em relação a afirmações verdadeiras): nem tudo o que é dito em um romance é falso (*Guerra e Paz*, de Tolstói, contém muitas declarações verdadeiras sobre a guerra franco-russa de Napoleão), nem todo texto que contenha afirmações falsas é considerado ficção (notícias falsas são apenas isso, não são ficções). A característica que define uma ficção é que os leitores devem **imaginar** os eventos e personagens descritos, não que eles sejam falsos (FRIGG, 2010a; SALIS, no prelo).

Giere (1988) defendeu a visão de que modelos "não físicos" são entidades abstratas. No entanto, há pouco acordo sobre a natureza dos objetos abstratos. Hale (1988, p. 86-87) lista não menos que doze diferentes caracterizações possíveis. Em publicações recentes, Thomasson (2020) e Thomson-Jones (2020) desenvolveram o que eles chamam de uma "visão artefatualista" de modelos, que é baseada na teoria de artefatos abstratos de Thomasson (1999). Essa visão concorda com a teoria do faz de conta de que o conteúdo do texto que apresenta um personagem ou modelo ficcional deve ser entendido como ocorrendo na imaginação, mas ao mesmo tempo insiste que, ao produzir tais descrições, os autores criam artefatos culturais abstratos que existem independentemente do autor ou dos leitores. O artefatualismo concorda com o platonismo que os objetos abstratos existem, mas insiste, contra o platonismo, que os objetos abstratos são trazidos à existência por meio de um ato criativo e não são eternos. Isso permite que o artefatualista preserve as vantagens da teoria do faz de conta e, ao mesmo tempo, mantém a visão realista de que personagens e modelos ficcionais realmente existem.

2.3 Estruturas de teoria dos conjuntos

Um ponto de vista influente considera os modelos como estruturas de teoria dos conjuntos. Esta posição pode ser rastreada até Suppes (1960) e agora, com ligeiras variantes, é sustentada pela maioria dos proponentes da chamada visão semântica das teorias. Existem diferenças entre as versões da visão semântica, mas com exceção de Giere (1988), todas as versões concordam que os modelos

são estruturas de um tipo ou de outro (DA COSTA; FRENCH, 2000).

Essa visão de modelos foi criticada por vários motivos. Uma crítica generalizada é que muitos tipos de modelos que desempenham um papel importante na ciência não são estruturas e não podem ser acomodados dentro da visão estruturalista dos modelos, que não pode explicar como esses modelos são construídos nem como eles funcionam no contexto da investigação (CARTWRIGHT, 1999; DOWNES, 1992; MORRISON, 1999). Exemplos de tais modelos são modelos interpretativos e modelos de mediação, discutidos posteriormente na Seção 4.2. Outra acusação contra a abordagem de teoria dos conjuntos é que as estruturas de teoria dos conjuntos por si mesmas não podem ser modelos representacionais – pelo menos se isso exige que compartilhem alguma estrutura com o alvo – porque a atribuição de uma estrutura a um sistema-alvo que faz parte de o mundo físico depende de uma descrição substantiva (não estrutural) do alvo, que vai além do que a abordagem estruturalista pode oferecer (NGUYEN; FRIGG, no prelo).

2.4 Descrições e equações

Uma posição consagrada diz que um modelo é uma descrição estilizada de um sistema-alvo. Argumentou-se que isso é o que os cientistas exibem em artigos e livros didáticos quando apresentam um modelo (ACHINSTEIN, 1968; BLACK, 1962). Esta visão não foi sujeita a críticas explícitas. No entanto, algumas das críticas que foram feitas contra a chamada visão sintática das teorias ameaçam igualmente uma compreensão linguística dos modelos. Primeiro, uma crítica padrão da visão sintática é que, ao associar uma teoria a uma formulação particular, a visão interpreta mal a identidade da teoria porque qualquer mudança na formulação resulta em uma nova teoria (SUPPE, 2000). Uma visão que associa modelos a descrições parece estar aberta às mesmas críticas. Em segundo lugar, os modelos têm propriedades diferentes das descrições: o modelo newtoniano do sistema solar consiste em esferas que se movem em órbitas, mas não faz sentido dizer isso sobre sua descrição. Por outro lado, as descrições têm propriedades que os modelos não possuem: uma descrição pode ser escrita em inglês e consistir em 517 palavras, mas o mesmo não pode ser dito de um modelo. Uma maneira de contornar essas dificuldades é associar o modelo ao conteúdo de uma descrição, e não à própria descrição. Para uma discussão de uma posição sobre modelos que se baseiam no conteúdo de uma descrição, consulte Salis (no prelo).

Uma versão contemporânea do descritivismo é a chamada visão de representação direta de Levy (2012, 2015) e Toon (2012). Essa visão compartilha com a visão ficcional dos modelos (Seção 2.2) a confiança na teoria do faz de conta de Walton, mas a usa de uma maneira diferente. A principal diferença é que as visões discutidas anteriormente veem a modelagem como a introdução de um veículo de representação, o modelo, que é distinto do alvo, e veem o problema como uma elucidação do tipo de coisa que o modelo é. Na visão de representação direta, não há modelos distintos do alvo; há apenas descrições de modelo e sistemas-alvo, sem modelos entre eles. Modelar, nesta visão, consiste em fornecer uma descrição imaginativa de coisas reais. Uma descrição de modelo prescreve imaginações sobre o sistema real; o pêndulo ideal, por exemplo, prescreve aos usuários do modelo que imaginem a mola real como perfeitamente elástica e o peso do pêndulo como uma massa pontual. Essa abordagem evita os problemas acima porque as condições de identidade dos modelos são dadas pelas condições dos jogos de faz de conta (e não pela sintaxe de uma descrição) e as atribuições de propriedade ocorrem na imaginação. Existem, no entanto, questões sobre como essa abordagem lida com modelos que não têm alvo (como modelos de éter ou das populações com quatro sexos) e sobre como os modelos assim entendidos lidam com idealizações. Para uma discussão desses pontos, consulte Frigg e Nguyen (2016), Poznic (2016) e Salis (no prelo).

Uma abordagem intimamente relacionada entende os modelos como equações. Esta é uma versão da visão de que os modelos são descrições, porque as equações são itens sintáticos que descrevem uma estrutura matemática. Os problemas que esta visão enfrenta são semelhantes aos que já encontramos: primeiro, pode-se descrever a mesma situação usando diferentes tipos de coordenadas e, como resultado, obter diferentes equações, mas sem, portanto, também obter um modelo diferente. Em segundo lugar, o modelo e a equação têm propriedades diferentes. Um pêndulo contém uma corda sem massa, mas a equação que descreve seu movimento não; e uma equação pode ser heterogênea, em contraposição ao sistema que ela descreve. É uma questão em aberto se esses problemas podem ser evitados apelando-se para uma abordagem imaginativa.

3. Epistemologia: As funções cognitivas dos modelos

Uma das principais razões pelas quais os modelos desempenham um papel tão importante na ciência é que eles desempenham uma série de funções cognitivas. Por exemplo, modelos são veículos para aprender sobre o mundo. Partes significativas da investigação científica são realizadas na forma de modelos, e não na própria realidade, porque ao estudar um modelo podemos descobrir características e averiguar fatos sobre o sistema que o modelo representa: os modelos permitem o "raciocínio substitutivo" (SWOYER, 1991). Por exemplo, estudamos a natureza do átomo de hidrogênio, a dinâmica de uma população ou o comportamento de um polímero estudando seus respectivos modelos. Esta função cognitiva dos modelos tem sido amplamente reconhecida na literatura, e alguns até sugerem que os modelos dão origem a um novo estilo de raciocínio, o "raciocínio baseado em modelos", segundo o qual "as inferências são feitas por meio da criação de modelos e da sua manipulação, adaptação e avaliação" (NERSESSIAN, 2010, p. 12; MAGNANI; NERSESSIAN; THAGARD, 1999; MAGNANI; NERSESSIAN, 2002; MAGNANI; CASADIO, 2016).

3.1 Aprendendo sobre modelos

Aprender sobre um modelo acontece em duas situações: na construção do modelo e em sua manipulação (MORGAN, 1999). Não existem regras fixas ou receitas para a construção de modelos, portanto, a própria atividade de descobrir o que se encaixa, e como, oferece uma oportunidade de aprender sobre o modelo. Uma vez que o modelo é construído, não aprendemos sobre suas propriedades olhando para ele; temos que usar e manipular o modelo para extrair seus segredos.

Dependendo do tipo de modelo que estamos lidando, construir e manipular um modelo envolve diferentes atividades que demandam diferentes metodologias. Os modelos físicos parecem ser diretos porque são usados em contextos experimentais comuns (por exemplo, colocamos o modelo de um carro no túnel de vento e medimos sua resistência ao ar). Portanto, no que diz respeito ao aprendizado sobre o modelo, os modelos físicos não dão origem a questões que vão além de questões relativas à experimentação em geral.

Isso é diferente com modelos ficcionais e abstratos. Que restrições existem para a construção de modelos ficcionais e abstratos, e como os manipulamos? Uma

resposta natural parece ser que fazemos isso através de um experimento mental. Diferentes autores (BROWN, 1991; GENDLER, 2000; NORTON, 1991; REISS, 2003; SORENSEN, 1992) exploraram essa linha de argumento, mas chegaram a conclusões muito diferentes e muitas vezes conflitantes sobre como os experimentos mentais são realizados e qual é o status de seus resultados.

Uma classe importante de modelos é de natureza computacional. Para alguns modelos é possível derivar resultados ou resolver equações de um modelo matemático analiticamente. Mas, muitas vezes, esse não é o caso. É neste ponto que os computadores têm um grande impacto, porque nos permitem resolver problemas que de outra forma seriam intratáveis. Consequentemente, os métodos computacionais nos fornecem conhecimento sobre (as consequências de) um modelo onde os métodos analíticos se omitem. Uma grande parte das pesquisas atuais em ciências naturais e sociais depende de simulações de computador, que ajudam os cientistas a explorar as consequências dos modelos que não podem ser investigados de outra forma. A formação e o desenvolvimento de estrelas e galáxias, a dinâmica das reações de íons pesados de alta energia, a evolução da vida, o surgimento de guerras, a progressão de uma economia, o comportamento moral e as consequências dos procedimentos de decisão em uma organização são explorados com simulações de computador, para citar apenas alguns exemplos.

Simulações de computador também são heuristicamente importantes. Eles podem sugerir novas teorias, modelos e hipóteses, por exemplo, com base em uma exploração sistemática do espaço de parâmetros de um modelo (HARTMANN, 1996). Mas as simulações de computador também apresentam riscos metodológicos. Por exemplo, eles podem fornecer resultados enganosos porque, devido à natureza discreta dos cálculos realizados em um computador digital, eles só permitem a exploração de uma parte do espaço completo de parâmetros, e este subespaço pode não refletir todas as características importantes do modelo. A gravidade desse problema é de certa forma atenuada pelo poder cada vez maior dos computadores modernos. Mas a disponibilidade de mais poder computacional também pode ter efeitos adversos: ela pode encorajar os cientistas a rapidamente apresentar modelos cada vez mais complexos, mas conceitualmente prematuros, envolvendo suposições ou mecanismos mal compreendidos e muitos parâmetros ajustáveis adicionais. Para uma discussão de um problema relacionado nas ciências sociais, confira Braun e Saam, 2015, cap. 3). Isso pode levar a um aumento na adequação empírica – o que pode ser bem-vindo para certas tarefas de previsão – mas não necessariamente leva a uma melhor compreensão dos mecanismos subjacentes. Como resultado,

o uso de simulações de computador pode alterar o peso que atribuímos aos vários objetivos da ciência. Finalmente, a disponibilidade do poder de computação pode seduzir os cientistas a fazer cálculos que não têm o grau de confiabilidade que se esperaria que tivessem. Isso acontece, por exemplo, quando computadores são usados para propagar distribuições de probabilidade no futuro, o que pode ser enganoso (FRIGG *et al*., 2014). Portanto, é importante não se deixar levar pelos meios que os novos e poderosos computadores oferecem e perder de vista os objetivos reais da pesquisa. Para uma discussão de outras questões relacionadas às simulações de computador, remetemos o leitor ao verbete sobre Simulações Computacionais na Ciência.

3.2 Aprendendo sobre sistemas-alvo

Assim que tivermos conhecimento sobre o modelo, esse conhecimento deve ser "traduzido" em conhecimento sobre o sistema-alvo. É neste ponto que a função representacional dos modelos se torna importante novamente: se um modelo representa, então ele pode nos instruir sobre a realidade, porque (pelo menos algumas das) partes ou aspectos do modelo têm partes ou aspectos correspondentes no mundo. Mas se a aprendizagem está conectada à representação, e se existem diferentes tipos de representações (analogias, idealizações, etc.), então também existem diferentes tipos de aprendizagem. Se, por exemplo, temos um modelo que consideramos ser uma representação realista, a transferência de conhecimento do modelo para o alvo é realizada de maneira diferente de quando lidamos com um análogo, ou um modelo que envolve suposições idealizadas. Para uma discussão das diferentes maneiras pelas quais a função representacional dos modelos pode ser explorada para aprender sobre o alvo, recomendamos ao leitor o verbete Representação Científica.

3.3 Explicando com modelos

Alguns modelos explicam. Mas como podem cumprir essa função, visto que costumam envolver idealizações? Esses modelos explicam apesar ou por causa das idealizações que envolvem? O uso explicativo de modelos pressupõe que eles representam, ou os modelos não representacionais também podem explicar? E que

tipo de explicação os modelos fornecem?

Há uma longa tradição que exige que o *explanans* de uma explicação científica seja verdadeiro. Podemos encontrar esse requisito no modelo nomológico-dedutivo (HEMPEL, 1965), bem como na literatura mais recente. Por exemplo, Strevens (2008, p. 297) afirma que "nenhuma abordagem causal da explicação [...] permite que modelos que não são verdadeiros expliquem". Para mais discussões, consulte também Colombo *et al*. (2015).

Autores que trabalham com essa tradição negam que as idealizações contribuam positivamente para a explicação e exploram como os modelos podem explicar apesar de serem idealizados. McMullin (1968, 1985) argumenta que uma explicação causal baseada em um modelo idealizado deixa de fora apenas características que são irrelevantes para a respectiva tarefa explicativa (ver também Salmon 1984 e Piccinini e Craver 2011 para uma discussão de esboços de mecanismos). Friedman (1974) argumenta que, na concepção unificacionista, um modelo mais realista (e, portanto, menos idealizado) é aquele que explica melhor. A ideia é que as idealizações podem, pelo menos em princípio, ser desidealizadas. Para uma discussão crítica desta afirmação, no contexto do debate sobre as explicações científicas, consulte Batterman (2002), Bokulich (2011), Morrison (2005, 2009), Jebeile e Kennedy (2015) e Rice (2015). Strevens (2008) argumenta que um modelo causal explicativo deve fornecer uma representação precisa das relações causais ou processos relevantes que o modelo compartilha com o sistema-alvo. As suposições idealizadas de um modelo não fazem diferença para o fenômeno em consideração e são, portanto, irrelevantes para a explicação. Em contraste, Potochnik (2017) e Rice (2015) argumentam que os modelos que explicam podem **distorcer** diretamente muitas causas que são relevantes [*difference-making causes*].

De acordo com a teoria de Woodward (2003), os modelos são ferramentas para descobrir sobre as relações causais que existem entre certos fatos ou processos, e são essas relações que fazem o trabalho explicativo. Mais especificamente, as explicações fornecem informações sobre os padrões de dependência contrafactual entre o *explanans* e o *explanandum* que

> nos permite ver que tipo de diferença teria feito para o *explanandum* se os fatores citados no explanans tivessem sido diferentes de várias maneiras (WOODWARD, 2003, p. 11).

Concepções sobre explicações causais também levaram a várias afirmações sobre como os modelos idealizados podem fornecer explicações, explorando até que ponto a idealização permite a representação incorreta de fatores causais irrelevantes pelo modelo explicativo (ELGIN; SOBER, 2002; STREVENS, 2004, 2008; POTOCHNIK, 2007; WEISBERG, 2007, 2013). No entanto, ter as características causalmente relevantes em comum com os sistemas reais continua a desempenhar o papel essencial em mostrar como os modelos idealizados podem ser explicativos.

Mas é realmente a verdade do *explanans* que torna o modelo explicativo? Outros autores seguem uma linha mais radical e argumentam que os falsos modelos explicam, não *apesar* de sua falsidade, mas *por causa* de sua falsidade. Cartwright (1983, p. 44) afirma que "a verdade não explica muito". Em sua concepção de modelos como "simulacros da explicação", ela sugere que podemos explicar um fenômeno construindo um modelo que encaixa o fenômeno na estrutura básica de uma grande teoria (1983, cap. 8). Nessa concepção, o próprio modelo é a explicação que buscamos. Isso se encaixa bem com as intuições científicas básicas, mas nos deixa com algumas questões: qual noção de explicação está em ação (ELGIN; SOBER, 2002) e que função explicativa as idealizações desempenham nos modelos explicativos (RICE, 2018, 2019)? Wimsatt (2007, cap. 6) enfatiza o papel dos falsos modelos como meios para chegar a teorias verdadeiras. Batterman e Rice (2014) argumentam que os modelos explicam por que os detalhes que caracterizam sistemas específicos não importam para a explicação. Bokulich (2008, 2009, 2011, 2012) segue uma linha de raciocínio semelhante e entende o poder explicativo dos modelos como intimamente relacionado à sua natureza ficcional. Bokulich (2009) e Kennedy (2012) apresentam concepções não representacionais dos modelos explicativos (JEBEILE; KENNEDY, 2015). Reiss (2012) e Woody (2004) fornecem discussões gerais sobre a relação entre representação e explicação.

3.4 Compreensão com modelos

Muitos autores apontam que a compreensão é um dos objetivos centrais da ciência (DE REGT, 2017; ELGIN, 2017; KHALIFA, 2017; POTOCHNIK, 2017). Em alguns casos, queremos entender um determinado fenômeno (por exemplo, por que o céu é azul?); em outros casos, queremos entender uma teoria científica específica (por exemplo, a mecânica quântica) que esclarece um fenômeno em questão. Às vezes, ganhamos compreensão de um fenômeno ao compreender a

teoria ou modelo correspondente. Por exemplo, a teoria eletromagnética de Maxwell nos ajuda a entender por que o céu é azul. No entanto, não é consenso se a compreensão de um fenômeno sempre pressupõe uma compreensão da teoria correspondente (DE REGT, 2009, p. 26).

Embora existam muitas maneiras diferentes de obter compreensão, os modelos e a atividade de modelagem científica são de particular importância (DE REGT *et al*., 2009; MORRISON, 2009; POTOCHNIK, 2017; Rice 2016). Essa ideia não é nova e pode ser encontrada nas famosas “Aulas de Baltimore” [*Baltimore Lectures*] sobre Dinâmica Molecular e a Teoria das Ondas da Luz de Lorde Kelvin, em 1884. Ele afirmou que "o teste de que 'Nós entendemos ou não um determinado assunto da física?' é 'Podemos fazer um modelo mecânico disso?'" (KELVIN, 1987, p. 111); Confira, também, Bailer-Jones (2009, cap. 2) e de Regt (2017, cap. 6).

Mas por que os modelos desempenham um papel tão crucial na compreensão de um assunto? Elgin (2017) argumenta que isso não ocorre apesar de, mas porque, os modelos são literalmente falsos. Ela entende os modelos falsos como "falsidades felizes" que ocupam o centro das atenções na epistemologia da ciência e menciona o modelo do gás ideal na mecânica estatística e o modelo de Hardy-Weinberg na genética como exemplos de modelos literalmente falsos que são centrais para suas respectivas disciplinas. A compreensão é holística e diz respeito a um tópico, disciplina ou assunto, em vez de afirmações ou fatos isolados. Obter a compreensão de um contexto significa ter

> um compromisso epistêmico com um corpo de informações abrangente e sistematicamente vinculado que seja fundamentado em fatos, que seja devidamente responsivo a razões ou evidências e permita inferências, argumentos e talvez ações não triviais em relação ao tópico em que as informações pertencem (ELGIN, 2017, p. 44).

e os modelos podem desempenhar um papel crucial na busca desses compromissos epistêmicos. Para uma discussão da descrição de modelos e compreensão em Elgin, consulte Baumberger e Brun (2017) e Frigg e Nguyen (no prelo).

Elgin (2017), Lipton (2009) e Rice (2016) argumentam que os modelos podem ser usados para compreender independentemente de sua capacidade de fornecer uma explicação. Outros autores, entre eles Strevens (2008, 2013),

argumentam que a compreensão pressupõe uma explicação científica e que

> um indivíduo tem compreensão científica de um fenômeno apenas através de uma explicação científica correta desse fenômeno (STREVENS, 2013, p. 510). No entanto, confira Sullivan e Khalifa (2019).

Nessa concepção, o entendimento consiste em uma forma particular de acesso epistêmico de um cientista individual a uma explicação. Para Strevens, esse aspecto é "apreensão", enquanto para de Regt (2017) é "inteligibilidade". É importante notar que tanto Strevens quanto de Regt sustentam que tais aspectos "subjetivos" são um tópico válido para investigações em filosofia da ciência. Esta visão contrasta com a visão tradicional (HEMPEL, 1965) que os delega ao reino da psicologia. Veja Friedman (1974), Trout (2002) e Reutlinger *et al*. (2018) para mais discussões sobre compreensão.

3.5 Outras funções cognitivas

Além das funções já mencionadas, foi enfatizado de várias maneiras que os modelos desempenham uma série de outras funções cognitivas. Knuuttila (2005, 2011) argumenta que o valor epistêmico dos modelos não se limita à sua função representacional e desenvolve uma abordagem que entende os modelos como artefatos epistêmicos que nos permitem reunir conhecimento de diversas maneiras. Nersessian (1999, 2010) enfatiza o papel dos modelos analógicos na formação de conceitos e outros processos cognitivos. Hartmann (1995) e Leplin (1980) discutem os modelos como ferramentas para a construção de teorias e enfatizam seu valor heurístico e pedagógico. Epstein (2008) lista uma série de funções específicas dos modelos nas ciências sociais. Peschard (2011) investiga a forma como os modelos podem ser usados para construir outros modelos e gerar novos sistemas-alvo. E Isaac (2013) discute usos não explicativos de modelos que não dependem de suas capacidades representacionais.

4. Modelos e teoria

Uma questão importante diz respeito à relação entre modelos e teorias. Existe um extenso espectro de posições, desde modelos subordinados a teorias até modelos independentes de teorias.

4.1 Modelos como subsidiários de teorias

Para discutir a relação entre modelos e teorias na ciência, é útil recapitular brevemente as noções de um modelo e de uma teoria na lógica. Uma teoria é considerada um conjunto (geralmente fechado por dedução) de sentenças em uma linguagem formal. Um modelo é uma estrutura (no sentido introduzido na Seção 2.3) que torna todas as sentenças de uma teoria verdadeiras quando seus símbolos são interpretados como se referindo a objetos, relações ou funções de uma estrutura. A estrutura é um modelo da teoria no sentido de que é corretamente descrita pela teoria (BELL; MACHOVER, 1977; HODGES, 1997). Os modelos lógicos às vezes também são chamados de "modelos de teoria" para indicar que são interpretações de um sistema formal abstrato.

Os modelos científicos às vezes importam da lógica a ideia de serem a interpretação de um cálculo abstrato (HESSE, 1967). Isso é saliente na física, onde as leis gerais – como a equação geral do movimento de Newton – estão no cerne de uma teoria. Essas leis são aplicadas a um sistema particular – por exemplo, um pêndulo – escolhendo uma função de força especial, fazendo suposições sobre a distribuição de massa do pêndulo, etc. O modelo resultante é uma interpretação (ou realização) da lei geral.

É importante manter as noções de um modelo lógico e representacional separadas (THOMSON-JONES, 2006): eles são conceitos distintos. Algo pode ser um modelo lógico sem ser um modelo representacional e *vice-versa*. Isso, entretanto, não significa que algo não possa ser um modelo em ambos os sentidos ao mesmo tempo. Na verdade, como Hesse (1967) aponta, muitos modelos em ciência são modelos lógicos e representacionais. O modelo do movimento planetários de Newton é um caso em questão: o modelo, que consiste em duas esferas perfeitas homogêneas localizadas em um espaço vazio que se atraem gravitacionalmente, é simultaneamente um modelo lógico (porque torna os axiomas da mecânica newtoniana verdadeiros quando são interpretados se referindo ao modelo) e um modelo representacional

(porque representa o Sol e a Terra reais).

Existem duas concepções principais de teorias científicas, a chamada visão sintática das teorias e a chamada visão semântica das teorias. Em ambas as concepções, os modelos desempenham um papel subsidiário às teorias, embora de maneiras muito diferentes. A visão sintática de teorias retém as noções lógicas de um modelo e uma teoria. Ele constrói uma teoria como um conjunto de sentenças em um sistema lógico axiomatizado e um modelo como uma interpretação alternativa de um determinado cálculo (BRAITHWAITE, 1953; CAMPBELL, 1957; NAGEL, 1961; SPECTOR, 1965). Se, por exemplo, pegarmos a matemática usada na teoria cinética dos gases e reinterpretar os termos desse cálculo de uma forma que os faça se referir a bolas de bilhar, as bolas de bilhar serão um modelo da teoria cinética dos gases no sentido de que todas as sentenças da teoria são verdadeiras. O modelo deve ser algo com o qual estamos familiarizados e serve ao propósito de tornar um cálculo formal abstrato mais palpável. Uma dada teoria pode ter modelos diferentes, e o modelo que escolhemos depende tanto dos nossos objetivos quanto do nosso conhecimento prévio. Os defensores da visão sintática discordam sobre a importância dos modelos. Carnap e Hempel entendiam que os modelos servem apenas um propósito pedagógico ou estético e são, em última análise, dispensáveis porque todas as informações relevantes estão contidas na teoria (CARNAP, 1938; HEMPEL, 1965; BAILER-JONES, 1999). Nagel (1961) e Braithwaite (1953), por outro lado, enfatizam o papel heurístico dos modelos, e Schaffner (1969) afirma que os termos teóricos obtêm pelo menos parte de seu significado através modelos.

A visão semântica das teorias dispensa sentenças em um sistema lógico axiomatizado e interpreta uma teoria como uma família de modelos. Nesta visão, uma teoria é literalmente uma classe, agrupamento ou família de modelos – modelos são os blocos de construção dos quais as teorias científicas são feitas. Diferentes versões da visão semântica funcionam com diferentes noções de um modelo, mas, como observado na Seção 2.3, na visão semântica, os modelos são principalmente interpretados como estruturas de teorias dos conjuntos.

4.2 Modelos como independentes de teorias

Tanto na visão sintática quanto na visão semântica das teorias, os modelos são vistos como subordinados à teoria, não desempenhando nenhum papel fora do contexto de uma teoria. Esta visão dos modelos foi desafiada de várias maneiras,

com autores apontando que os modelos apresentam vários graus de liberdade da teoria e funcionam de forma autônoma em muitos contextos. A independência pode assumir diversas formas, e grande parte da literatura sobre modelos se preocupa com a investigação dessas várias formas de independência.

Modelos completamente independentes de teorias. O distanciamento mais radical da análise de modelos centrada na teoria é a compreensão de que existem modelos que são completamente independentes de qualquer teoria. Um exemplo de tal modelo é o modelo Lotka-Volterra. O modelo descreve a interação de duas populações: uma população de predadores e uma população de presas (WEISBERG, 2013). O modelo foi construído usando apenas suposições relativamente de senso comum sobre predadores e presas e matemática das equações diferenciais. Não houve apelo a uma teoria das interações presa-predador ou uma teoria do crescimento populacional, e o modelo é independente de teorias sobre o seu objeto de estudo. Se um modelo é construído em um domínio onde nenhuma teoria está disponível, então o modelo é algumas vezes referido como um "modelo substituto" (GROENEWOLD, 1961), porque o modelo substitui uma teoria.

Modelos como meio de explorar teorias. Os modelos também podem ser usados para explorar teorias (MORGAN; MORRISON, 1999). Uma maneira óbvia de isso acontecer é quando um modelo é um modelo lógico de uma teoria (consulte a Seção 4.1). Um modelo lógico é um conjunto de objetos e propriedades que tornam uma sentença formal verdadeira e, portanto, pode-se ver no modelo como os axiomas da teoria atuam em um ambiente específico e que tipos de comportamento eles ditam. Mas nem todos os modelos usados para explorar teorias são modelos lógicos e os modelos podem representar características de teorias de outras maneiras. Como exemplo, considere a Teoria do Caos. As equações dos sistemas não lineares, como as que descrevem o problema dos três corpos, têm soluções muito complexas para se estudar manualmente e até mesmo as simulações de computador são limitadas de várias maneiras. Considerações abstratas sobre o comportamento qualitativo das soluções mostram que existe um mecanismo que foi apelidado de "esticar e dobrar". Para obter uma ideia da complexidade da dinâmica estica-e-dobra, Smale propôs estudar um modelo simples de fluxo – agora conhecido como "mapa da ferradura" (TABOR, 1989) – que fornece importantes insights sobre a natureza do esticar e dobrar. Outros exemplos de modelos desse tipo são o modelo do anel de Kac que é usado para estudar propriedades de equilíbrio de sistemas em mecânica estatística (LAVIS, 2008) e a cúpula de Norton em mecânica newtoniana (NORTON, 2003).

Modelos como complementos de teorias. Uma teoria pode ser especificada de forma incompleta no sentido de que apenas impõe certas restrições gerais, mas se omite sobre os detalhes de situações concretas, que são fornecidas por um modelo (REDHEAD, 1980). Um caso especial dessa situação é quando uma teoria qualitativa é conhecida e o modelo introduz medidas quantitativas (APOSTEL, 1961). O exemplo de uma teoria que é subdeterminada desta forma apresentado por Redhead é a teoria quântica de campos axiomática, que apenas impõe certas restrições gerais aos campos quânticos, mas não fornece uma descrição de campos específicos. Harré (2004) observa que os modelos podem complementar as teorias ao fornecer mecanismos para processos que não são especificados na teoria, embora sejam responsáveis por ocasionar os fenômenos observados.

As teorias podem ser muito difíceis de abordar. Nesses casos, um modelo pode complementar uma teoria fornecendo uma versão simplificada do cenário teórico que permite uma solução. A cromodinâmica quântica, por exemplo, não pode ser usada facilmente para investigar a física de um núcleo atômico, embora seja a teoria fundamental relevante. Para contornar essa dificuldade, os físicos constroem modelos fenomenológicos tratáveis (como o modelo de sacola do MIT) que descrevem efetivamente os graus de liberdade relevantes do sistema em consideração (HARTMANN, 1999, 2001). A vantagem desses modelos é que eles produzem resultados onde as teorias se omitem. A desvantagem é que muitas vezes não fica claro como entender a relação entre o modelo e a teoria, já que os dois são, falando estritamente, contraditórios.

Modelos como teorias preliminares. A noção de um modelo como substituto de uma teoria está intimamente relacionada à noção de um **modelo de mudança científica**. Esse termo foi cunhado por Leplin (1980), que apontou como os modelos eram úteis no desenvolvimento da teoria quântica inicial, e agora é usado como um termo genérico que cobre casos em que os modelos são uma espécie de exercício preliminar à teoria.

Também intimamente relacionado está a noção de um **modelo de sondagem** (ou "modelo de estudo"). Modelos desse tipo não desempenham uma função representacional e não se espera que instruam sobre nada além do próprio modelo. O objetivo desses modelos é testar novas ferramentas teóricas que serão usadas posteriormente para construir modelos representacionais. Na Teoria de Campos, por exemplo, o chamado modelo φ^4 foi estudado extensivamente, não porque se acreditava que ele representasse algo real, mas porque servia várias funções heurísticas: a simplicidade do modelo φ^4 permitia aos físicos "ter uma idea" de como

seriam as teorias quânticas de campo e para extrair algumas características gerais que este modelo simples compartilhava com outros mais complicados. Os físicos podiam estudar técnicas complicadas, como a renormalização em um ambiente simples, e era possível se familiarizar com mecanismos importantes – neste caso a quebra de simetria – que poderiam ser usados posteriormente em diferentes contextos (HARTMANN, 1995). Isso não acontece apenas na física. Como Wimsatt (1987, 2007) aponta, um modelo falso em genética pode desempenhar muitas funções úteis, entre elas as seguintes: o modelo falso pode ajudar a responder a questões sobre modelos mais realistas, fornecer uma arena para responder a questões sobre propriedades de modelos mais complexos, "realçar" fenômenos que não seriam vistos de outra forma, servir como um caso limite de um modelo mais geral (ou dois modelos falsos podem definir os extremos de um espectro de casos em que o caso real supostamente se localiza), ou conduzir a identificação das variáveis relevantes e a estimativa dos seus valores.

Modelos interpretativos. Cartwright (1983, 1999) argumenta que os modelos não auxiliam apenas na aplicação de teorias que são de alguma forma incompletas; ela afirma que os modelos também estão envolvidos sempre que uma teoria com uma estrutura matemática abrangente é aplicada. As principais teorias da física – mecânica clássica, eletrodinâmica, mecânica quântica e assim por diante – se enquadram nessa categoria. Teorias desse tipo são formuladas em termos de conceitos abstratos que precisam ser concretizados para que a teoria forneça uma descrição do sistema-alvo e, concretizando os conceitos relevantes, objetos e processos idealizados são introduzidos. Por exemplo, ao aplicar a mecânica clássica, o conceito abstrato de força deve ser substituído por uma força concreta como a gravidade. Para obter equações tratáveis, este procedimento deve ser aplicado a um cenário simplificado, por exemplo, o de dois planetas perfeitamente esféricos e homogêneos em um espaço vazio, ao invés da realidade em sua complexidade total. O resultado é um modelo interpretativo, que fundamenta a aplicação de teorias matemáticas a sistemas-alvo no mundo real. Tais modelos são independentes da teoria no sentido de que a teoria não determina sua forma e, ainda assim, são necessários para a aplicação da teoria a um problema concreto.

Modelos como mediadores. A relação entre modelos e teorias pode ser complicada e desordenada. Os colaboradores de uma coleção programática de ensaios editada por Morgan e Morrison (1999) partem da ideia de que modelos são instrumentos de mediação entre as teorias e o mundo. Os modelos são "agentes autônomos" no sentido de que são independentes de ambas as teorias e de seus

sistemas-alvo, e é essa independência que os permite mediar os dois. As teorias não nos fornecem algoritmos para a construção de um modelo; elas não são "máquinas de venda automáticas" [*vending machines*] nas quais se insere um problema e um modelo simplesmente aparece (CARTWRIGHT, 1999). A construção de um modelo frequentemente requer conhecimento detalhado sobre materiais, esquemas de aproximação e sua configuração, e estes não são fornecidos pela teoria correspondente. Além disso, o funcionamento interno de um modelo é frequentemente conduzido por uma série de teorias diferentes trabalhando cooperativamente. Na modelagem climática contemporânea, por exemplo, elementos de diferentes teorias – entre elas dinâmica de fluidos, termodinâmica, eletromagnetismo – são colocados para trabalhar cooperativamente. O que produz os resultados não é a aplicação rigorosa de uma teoria, mas as vozes de diferentes teorias colocadas em coro em um modelo.

Em casos complexos, como o estudo de um sistema de laser ou do clima global, os modelos e as teorias podem ficar tão emaranhados que não fica claro onde deve ser traçada uma linha entre os dois: onde termina o modelo e começa a teoria? Este não é um problema apenas para a análise filosófica; ele também surge na prática científica. Bailer-Jones (2002) entrevistou um grupo de físicos sobre sua compreensão sobre modelos e sua relação com teorias, e relatou visões amplamente divergentes: (i) não há diferença substantiva entre modelo e teoria; (ii) os modelos **tornam-se** teorias quando seu grau de confirmação aumenta; (iii) os modelos contêm simplificações e omissões, enquanto as teorias são precisas e completas; (iv) as teorias são mais gerais do que os modelos, e a modelagem trata da aplicação de teorias gerais a casos específicos. A primeira sugestão parece muito radical para fazer justiça a muitos aspectos da prática, onde a distinção entre modelos e teorias é claramente realizada. A segunda visão está de acordo com a linguagem comum, onde os termos "modelo" e "teoria" às vezes são usados para expressar a atitude de alguém em relação a uma hipótese particular. A frase "é apenas um modelo" indica que a hipótese em jogo é afirmada apenas provisoriamente ou até mesmo conhecida como falsa, enquanto algo recebe o rótulo de "teoria" se tiver adquirido algum grau de aceitação geral. No entanto, este uso de "modelo" é diferente dos usos que vimos nas Seções 1 a 3 e, portanto, não tem utilidade se pretendemos compreender a relação entre modelos científicos e teorias (e, a propósito, pode-se igualmente descartar afirmações especulativas como sendo "apenas uma teoria"). A terceira proposta é correta ao associar modelos a idealizações e simplificações, mas extrapola ao restringir isso a modelos; na verdade, as teorias

também podem conter idealizações e simplificações. A quarta visão parece intimamente alinhada com os modelos interpretativos e a ideia de que os modelos são mediadores, mas ser mais geral é uma noção gradual e, portanto, não fornece um critério bem definido para distinguir entre teorias e modelos.

5. Modelos e outros debates na filosofia da ciência

O debate sobre modelos científicos tem repercussões importantes para outras questões na filosofia da ciência. Para um relato histórico da discussão filosófica sobre modelos, confira Bailer-Jones (1999). Tradicionalmente, os debates sobre, digamos, realismo científico, reducionismo e leis da natureza eram expressos em termos de teorias, porque as teorias eram vistas como os principais portadores do conhecimento científico. Uma vez que os modelos são reconhecidos como ocupando um lugar importante na ciência, essas questões devem ser reconsideradas com foco nos modelos. A questão é se, e se sim como, as discussões dessas questões mudam quando mudamos o foco das teorias para os modelos. Até agora, nenhum relato abrangente baseado em modelos de qualquer uma dessas questões surgiu, mas os modelos deixaram marcas importantes nas discussões desses tópicos.

5.1 Modelos, realismo e leis da natureza

Como vimos na Seção 1, os modelos geralmente fornecem uma representação distorcida de seus alvos. Se alguém entende a ciência essencialmente baseada em modelos, isso pode ser interpretado como uma interpretação antirrealista da ciência. Os realistas, no entanto, negam que a presença de idealizações nos modelos torne uma abordagem realista da ciência impossível e apontam que um bom modelo, embora não seja literalmente verdadeiro, geralmente é pelo menos aproximadamente verdadeiro, e/ou pode ser melhorado por desidealização (LAYMON, 1985; MCMULLIN, 1985; NOWAK, 1979; BRZEZINSKI; NOWAK, 1992).

Além das frequentes preocupações sobre a falta de clareza da noção de verdade aproximada os antirrealistas questionaram essa resposta por duas razões (relacionadas). Em primeiro lugar, como Cartwright (1989) aponta, não há razão para supor que sempre se pode melhorar um modelo adicionando correções desidealizantes. Em segundo lugar, a desidealização parece não estar de acordo

com a prática científica, porque é incomum ver cientistas investindo em desidealizar um modelo existente. Em vez disso, eles mudam para uma estrutura de modelagem diferente, uma vez que os ajustes a serem feitos ficam muito complicados (HARTMANN, 1998). Os vários modelos do núcleo atômico são um exemplo: uma vez que foi percebido que os efeitos de camada são importantes para entender vários fenômenos subatômicos, o modelo da gota líquida (coletivo) foi colocado de lado e o modelo de camadas (de partícula única) foi desenvolvido para dar conta das descobertas correspondentes. Outra dificuldade com a desidealização é que a maioria das idealizações não é "controlada". Por exemplo, não está claro de que maneira alguém poderia desidealizar o modelo da sacola do MIT para eventualmente chegar à cromodinâmica quântica, a teoria supostamente correta subjacente a ele.

Um outro argumento antirrealista, o "argumento dos modelos incompatíveis", toma como ponto de partida a observação de que os cientistas muitas vezes usam com sucesso vários modelos incompatíveis de um mesmo sistema-alvo para fins preditivos (MORRISON, 2000). Esses modelos aparentemente se contradizem, pois atribuem propriedades diferentes ao mesmo sistema-alvo. Na física nuclear, por exemplo, o modelo da gota líquida explora a analogia do núcleo atômico com uma gota de fluido (carregada), enquanto o modelo de camadas descreve propriedades nucleares em termos de propriedades de prótons e nêutrons, os constituintes de um núcleo atômico. Essa prática parece causar um problema para o realismo científico: os realistas normalmente sustentam que há uma conexão estreita entre o sucesso preditivo de uma teoria e o fato de ela ser, pelo menos aproximadamente, verdadeira. Mas se vários modelos do mesmo sistema são bem-sucedidos em suas predições e se esses modelos são mutuamente inconsistentes, é difícil sustentar que todos são aproximadamente verdadeiros.

Os realistas podem reagir a esse argumento de várias maneiras. Em primeiro lugar, eles podem contestar a afirmação de que os modelos em questão são de fato bem-sucedidos em suas predições. Se os modelos não forem bons preditores, o argumento será descreditado. Em segundo lugar, eles podem defender uma versão de "realismo de perspectiva" (GIERE, 2006; MASSIMI, 2017; RUEGER, 2005). Os defensores desta posição (que às vezes também é chamada de "perspectivismo") a situam em algum lugar entre o realismo científico "padrão" e o antirrealismo, mas exatamente onde se encontra sua posição é assunto de um debate ativo (MASSIMI, 2018a, 2018b; SAATSI, 2016; TELLER, 2018). E confira, também, as contribuições em Massimi e McCoy (2019). Em terceiro lugar, os realistas podem negar que haja um problema de antemão, porque os modelos científicos,

que são sempre idealizados e, portanto, estritamente falsos, são apenas o veículo errado para fazer qualquer observação sobre a realidade (que deve ser discutida em termos de teorias).

Um foco particular do debate sobre o realismo são as leis da natureza, no qual surgem as questões sobre o que são as leis e se elas são refletidas de forma verdadeira em nossas representações científicas. De acordo com as duas concepções atualmente dominantes, a abordagem dos melhores sistemas e a abordagem necessitarista, as leis da natureza são entendidas como de escopo universal, o que significa que se aplicam a tudo o que existe no mundo. Essa visão das leis não parece se encaixar bem com uma visão que coloca os modelos no centro da pesquisa científica. Qual é o papel das leis gerais na ciência se são os modelos que representam o que está acontecendo no mundo? E como os modelos e as leis estão relacionados?

Uma resposta possível a essas perguntas é argumentar que as leis da natureza governam entidades e processos em um modelo, e não no mundo. As leis fundamentais, nessa abordagem, não declaram fatos sobre o mundo, mas são verdadeiras para entidades e processos no modelo. Essa visão foi defendida de diversas maneiras: Cartwright (1983) argumenta que todas as leis são leis *ceteris paribus*. Cartwright (1999) faz uso de "capacidades" (que ela considera anteriores às leis) e introduz a noção de uma "máquina nomológica". Isto é

> um arranjo fixo (suficiente) de componentes, ou fatores, com capacidades estáveis (suficientes) que no tipo certo de ambiente estável (suficiente) irá, repetidamente, dar origem ao tipo de comportamento regular que representamos em nossas leis científicas (1999, p. 50).

Giere (1999) argumenta que as leis de uma teoria são melhores entendidas, não como codificando verdades gerais sobre o mundo, mas sim como afirmações incompletas que podem ser preenchidas de várias maneiras no processo de construção de modelos científicos mais específicos. Posições semelhantes também foram defendidas por Teller (2001) e van Fraassen (1989).

5.2 Modelos e reducionismo

O problema dos modelos múltiplos mencionado na Seção 5.1 também levanta a questão de como os diferentes modelos estão relacionados. Evidentemente, vários modelos para o mesmo sistema-alvo geralmente não têm relação dedutiva, pois muitas vezes se contradizem. Alguns (CARTWRIGHT, 1999; HACKING, 1983) sugeriram uma ideia da ciência segundo a qual não existem relações sistemáticas entre os diferentes modelos. Alguns modelos são interligados porque representam o mesmo sistema-alvo, mas isso não significa que eles teriam qualquer outra forma de relacionamento (dedutivo ou outro). Somos confrontados com uma colcha de retalhos de modelos, todos os quais são mantidos *ceteris paribus* em seus domínios específicos de aplicabilidade.

Alguns argumentam que essa imagem é, pelo menos parcialmente, incorreta porque existem várias relações interessantes que existem entre diferentes modelos ou teorias. Essas relações variam de relações reducionistas completas (SCHEIBE, 1997, 1999, 2001: esp. cap. V.23 e V.24) e aproximações controladas sobre relações de limite singulares (BATTERMAN, 2016) a relações estruturais (GÄHDE, 1997) e relações frouxas chamadas "estórias" (HARTMANN, 1999; BOKULICH, 2003; TELLER, 2002; HARTMANN *et al*., 2008). Essas sugestões foram feitas com base em estudos de caso, e resta entender se uma descrição mais geral dessas relações pode ser fornecida e se uma justificativa mais profunda para elas pode ser fornecida, por exemplo, dentro de uma estrutura bayesiana. Os primeiros passos em direção a uma compreensão bayesiana das relações redutivas pode ser encontrada em Dizadji-Bahmani *et al*. (2011), em Liefke e Hartmann (2018) e Tešić (2019).

Os modelos também figuram no debate sobre redução e emergência na física. Aqui, alguns autores argumentam que a abordagem moderna de renormalização desafia o modelo reducionista de Nagel (1961) ou a doutrina mais ampla sobre reduções. Para uma discussão crítica, ver, por exemplo, Batterman (2002, 2010, 2011), Morrison (2012) e Saatsi e Reutlinger (2018). Dizadji-Bahmani *et al*. (2010) fornecem uma defesa do modelo reducionista de Nagel-Schaffner e Butterfield (2011a, b, 2014) argumenta que a renormalização é consistente com o reducionismo de Nagel. Palacios (2019) mostra que as transições de fase são compatíveis com o reducionismo, e Hartmann (2001) argumenta que o programa de pesquisa de teorias de campo efetivo é consistente com o reducionismo (BAIN, 2013; FRANKLIN, no prelo). Rosaler (2015) defende uma forma "local" de redução que entende a relação fundamental de redução mantida entre modelos, não entre teorias, o que

é, no entanto, compatível com o modelo reducionista de Nagel-Schaffner.

Nas ciências sociais, modelos baseados em agentes [*agent-based models* – ABMs] são cada vez mais usados (KLEIN *et al*., 2018). Esses modelos mostram como padrões comportamentais surpreendentemente complexos na escala macro podem emergir de um pequeno número de regras comportamentais simples para os agentes individuais e suas interações. Isso levanta questões semelhantes às questões mencionadas acima sobre redução e emergência na física, mas até agora só encontramos comentários esparsos sobre redução na literatura. Consulte Weisberg e Muldoon (2009) e Zollman (2007) para a aplicação dos ABM na epistemologia e na estrutura social da ciência, e Colyvan (2013) para uma discussão de questões metodológicas levantadas por modelos normativos em geral.

Referências

ACHINSTEIN, P. **Concepts of Science**: A Philosophical Analysis. Baltimore: Johns Hopkins Press, 1968.

AKERLOF, G. A. The Market for Lemons: Quality Uncertainty and the Market Mechanism. **The Quarterly Journal of Economics**, v. 84, n. 3, p. 488-500, 1970. DOI:10.2307/1879431.

APOSTEL, L. Towards the Formal Study of Models in the Non-Formal Sciences. *In:* FREUDENTHAL, H. (ed.). **The Concept and the Role of the Model in Mathematics and Natural and Social Sciences**. Dordrecht: Reidel, 1961, p. 1-37. DOI: 10.1007/978-94-010-3667-2_1.

BAILER-JONES, D. M. Tracing the Development of Models in the Philosophy of Science. *In:* MAGNANI, L.; NERSESSIAN, N. J.; THAGARD, P. (ed.). **Model-Based Reasoning in Scientific Discovery.** Boston: Springer US, 1999, p. 23-40. DOI: 10.1007/978-1-4615-4813-3_2.

BAILER-JONES, D. M. Scientists' Thoughts on Scientific Models. **Perspectives on Science**, v. 10, n. 3, p. 275-301, 2002. DOI:10.1162/106361402321899069.

BAILER-JONES, D. M. **Scientific Models in Philosophy of Science.** Pittsburgh: University of Pittsburgh Press, 2009.

BAILER-JONES, D. M.; BAILER-JONES, C. A. L. Modeling Data: Analogies in Neural Networks, Simulated Annealing and Genetic Algorithms. *In:* MAGNANI, L.; NERSESSIAN, N. J. (ed.). **Model-Based Reasoning**: Science, Technology, Values. Boston: Springer US, 2002, p.147-165. DOI:10.1007/978-1-4615-0605-8_9.

BAIN, J. Emergence in Effective Field Theories. **European Journal for Philosophy of Science**, v. 3, n. 3, p. 257-273, 2013. DOI:10.1007/s13194-013-0067-0.

BANDYOPADHYAY, P. S.; FORSTER, M. R. (eds.). **Philosophy of Statistics** (Handbook of the Philosophy of Science 7). Amsterdam: Elsevier, 2011.

BARBEROUSSE, A.; LUDWIG, P. Fictions and Models. *In:* SUÁREZ, M. (ed.). **Fictions in Science**: Philosophical Essays on Modeling and Idealization. London: Routledge, 2009, p. 56-75. DOI:10.4324/9780203890103.

BARTHA, P. **By Parallel Reasoning**: The Construction and Evaluation of Analogical Arguments. New York: Oxford University Press, 2010. DOI: 10.1093/acprof:oso/9780195325539.001.0001.

BARTHA, P. Analogy and Analogical Reasoning. *In:* ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy**, 2019. Disponível em: https://plato.stanford.edu/archives/spr2019/entries/reasoning-analogy/. Acesso em: 06 out. 2021.

BATTERMAN, R. W. **The Devil in the Details**: Asymptotic Reasoning in Explanation, Reduction, and Emergence. Oxford: Oxford University Press, 2002. DOI:10.1093/0195146476.001.0001.

BATTERMAN, R. W. Intertheory Relations in Physics. *In:* ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy,** 2016. Disponível em: https://plato.stanford.edu/archives/fall2016/entries/physics-interrelate. Acesso em: 06 out. 2021.

BATTERMAN, R. W. Reduction and Renormalization. *In:* ERNST, G.; HÜTTEMANN, A. (ed.). **Time, Chance and Reduction**: Philosophical Aspects of Statistical Mechanics. Cambridge: Cambridge University Press, 2010, p. 159-179. 2010

BATTERMAN, R. W. Emergence, Singularities, and Symmetry Breaking. **Foundations of Physics**, v. 41, n. 6, p. 1031-1050, 2011. DOI:10.1007/s10701-010-9493-4.

BATTERMAN, R. W.; RICE, C. C. Minimal Model Explanations. **Philosophy of Science**, v. 81, n. 3, p. 349-376, 2014. DOI:10.1086/676677.

BAUMBERGER, C.; BRUN, G. Dimensions of Objectual Understanding. *In:* GRIMM, S. R.; BAUMBERGER, C.; AMMON, S. (ed.). **Explaining Understanding**: New Perspectives from Epistemology and Philosophy of Science. New York: Routledge, 2017, p. 165-189.

BELL, J.; MACHOVER, M. **A Course in Mathematical Logic.** Amsterdam: North-Holland, 1977.

BERRY, M. Singular Limits. **Physics Today**, v. 55, n. 5, p. 10-11, 2002. DOI:10.1063/1.1485555.

BLACK, M. **Models and Metaphors**: Studies in Language and Philosophy. Ithaca: Cornell University Press, 1962.

BOGEN, J.; WOODWARD, J. Saving the Phenomena. **The Philosophical Review**, v. 97, n. 3, p. 303-352, 1988. DOI:10.2307/2185445.

BOKULICH, A. Horizontal Models: From Bakers to Cats. **Philosophy of Science**, v. 70, n. 3, p. 609-627, 2003. DOI:10.1086/376927.

BOKULICH, A. **Reexamining the Quantum–Classical Relation**: Beyond Reductionism and Pluralism. Cambridge: Cambridge University Press, 2008. DOI:10.1017/CBO9780511751813.

BOKULICH, A. Explanatory Fictions. SUÁREZ, M. (ed.). **Fictions in Science**: Philosophical Essays on Modeling and Idealization. London: Routledge, 2009, p. 91-109.

BOKULICH, A. How Scientific Models Can Explain. **Synthese**, v. 180, n. 1, p. 33-45, 2011. DOI:10.1007/s11229-009-9565-1.

BOKULICH, A. Distinguishing Explanatory from Nonexplanatory Fictions. **Philosophy of Science**, v. 79, n. 5, p. 725-737, 2012. DOI:10.1086/667991.

BRAITHWAITE, R. **Scientific Explanation.** Cambridge: Cambridge University Press, 1953.

BRAUN, N.; SAAM, N. J. (ed.). **Handbuch Modellbildung und Simulation in den Sozialwissenschaften**. Wiesbaden: Springer Fachmedien, 2015. DOI:10.1007/978-3-658-01164-2.

BREWER, W. F.; CHINN, C. A. Scientists Responses to Anomalous Data: Evidence from Psychology, History, and Philosophy of Science. **Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association**, v. 1, p. 304-313, 1994. DOI:10.1086/psaprocbienmeetp.1994.1.193035.

BROWN, J. **The Laboratory of the Mind**: Thought Experiments in the Natural Sciences. London: Routledge, 1991.

BRZEZINSKI, J.; NOWAK, L. (ed.). **Idealization III**: Approximation and Truth. Amsterdam: Rodopi, 1992.

BUTTERFIELD, J. Emergence, Reduction and Supervenience: A Varied Landscape. **Foundations of Physics**, v. 41, n. 6, p. 920-959, 2011a. DOI:10.1007/s10701-011-9549-0.

BUTTERFIELD, J. Less Is Different: Emergence and Reduction Reconciled", **Foundations of Physics**, v. 41, n. 6, p. 1065-1135, 2011b. DOI:10.1007/s10701-010-9516-1.

BUTTERFIELD, J. Reduction, Emergence, and Renormalization. **Journal of Philosophy**, v. 111, n. 1, p. 5-49, 2014. DOI:10.5840/jphil201411111.

CALLENDER, C.; COHEN, J. There Is No Special Problem about Scientific Representation. **Theoria**, v. 55, n. 1, p. 67-85, 2006.

CAMPBELL, N. **Physics**: The Elements. Cambridge: Cambridge University Press, 1957.

CARNAP, R. Foundations of Logic and Mathematics. *In:* NEURATH, O.; MORRIS, C.; CARNAP, R. (ed.). **International Encyclopaedia of Unified Science.** v. 1. Chicago: University of Chicago Press, 1938, p. 139-213.

CARTWRIGHT, N. **How the Laws of Physics Lie.** Oxford: Oxford University, Press, 1983. DOI:10.1093/0198247044.001.0001.

CARTWRIGHT, N. **Nature's Capacities and Their Measurement**. Oxford: Oxford University Press, 1989. DOI:10.1093/0198235070.001.0001.

CARTWRIGHT, N. **The Dappled World**: A Study of the Boundaries of Science. Cambridge: Cambridge University Press, 1999. DOI:10.1017/CBO9781139167093.

COLOMBO, M.; HARTMANN, S.; VAN IERSEL, R. Models, Mechanisms, and Coherence. **The British Journal for the Philosophy of Science**, v. 66, n. 1, p. 181-212, 2015. DOI:10.1093/bjps/axt043.

COLYVAN, M. Idealisations in Normative Models. **Synthese**, v. 190, n. 8, p. 1337-1350, 2013. DOI:10.1007/s11229-012-0166-z.

CONTESSA, G. Scientific Models and Fictional Objects. **Synthese**, v. 172, n. 2, p. 215-229, 2010. DOI:10.1007/s11229-009-9503-2.

CROWTHER, K.; LINNEMANN, N. S.; WÜTHRICH, C.; What We Cannot Learn from Analogue Experiments", **Synthese**, 2019. DOI:10.1007/s11229-019-02190-0.

DA COSTA, N.; FRENCH, S. Models, Theories, and Structures: Thirty Years On. **Philosophy of Science**, v. 67 sup, p.S116-S127, 2000. DOI:10.1086/392813.

DARDASHTI, R.; HARTMANN, S.; THÉBAULT, K.; WINSBERG, E. Hawking Radiation and Analogue Experiments: A Bayesian Analysis. **Studies in History and Philosophy of Science Part B**: Studies in History and Philosophy of Modern Physics, v. 67, p. 1-11, 2019. DOI:10.1016/j.shpsb.2019.04.004.

DARDASHTI, R.; THÉBAULT, K. P. Y.; WINSBERG, E. Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity. **The British Journal for the Philosophy of Science**, v. 68, n. 1, p. 55-89, 2017. DOI:10.1093/bjps/axv010.

DE REGT, H. Understanding and Scientific Explanation. DE REGT, H.; LEONELLI, S.; EIGNER, K. (ed.). **Scientific Understanding**: Philosophical Perspectives. Pittsburgh: University of Pittsburgh Press, 2009, p. 21-42.

DE REGT, H. **Understanding Scientific Understanding**. Oxford: Oxford University Press, 2017. DOI:10.1093/oso/9780190652913.001.0001.

DE REGT, H.; LEONELLI, S.; EIGNER, K. (ed.). **Scientific Understanding**: Philosophical Perspectives. Pittsburgh: University of Pittsburgh Press, 2009.

DIZADJI-BAHMANI, F.; FRIGG, R.; HARTMANN, S. Who's Afraid of Nagelian Reduction? **Erkenntnis**, v. 73, n. 3, p. 393-412, 2010. DOI:10.1007/s10670-010-9239-x.

DIZADJI-BAHMANI, F.; FRIGG, R.; HARTMANN, S. Confirmation and Reduction: A Bayesian Account. **Synthese**, v. 179, n. 2, p. 321-338, 2011. DOI:10.1007/s11229-010-9775-6.

DOWNES, S. M. The Importance of Models in Theorizing: A Deflationary Semantic View. **Proceedings of the Biennial Meeting of the Philosophy of Science Association**, v. 1, p. 142-153, 1992. DOI:10.1086/psaprocbienmeetp.1992.1.192750.

ELGIN, C. Z. Telling Instances. *In:* FRIGG, R.; HUNTER, M. (ed.). **Beyond Mimesis and Convention** (Boston Studies in the Philosophy of Science 262). Dordrecht: Springer Netherlands, 2010. p. 1-17, 2010. DOI:10.1007/978-90-481-3851-7_1.

ELGIN, C. Z. **True Enough**. Cambridge: MIT Press, 2017.

ELGIN, M.; SOBER, E. Cartwright on Explanation and Idealization. **Erkenntnis**, v. 57, n. 3, p. 441-450, 2002. DOI:10.1023/A:1021502932490.

EPSTEIN, J. M. Why Model?. **Journal of Artificial Societies and Social Simulation**, v. 11, n. 4, p. 12, 2008.

FISHER, G. The Autonomy of Models and Explanation: Anomalous Molecular Rearrangements in Early Twentieth-Century Physical Organic Chemistry. **Studies in History and Philosophy of Science Part A**, v. 37, n. 4, p. 562-584, 2006. DOI:10.1016/j.shpsa.2006.09.009.

FRANKLIN, A. Whence the Effectiveness of Effective Field Theories? **The British Journal for the Philosophy of Science**, 2018. DOI:10.1093/bjps/axy050.

FREUDENTHAL, H. (ed.). **The Concept and the Role of the Model in Mathematics and Natural and Social Sciences**. Dordrecht: Reidel, 1961. DOI:10.1007/978-94-010-3667-2.

FRIEDMAN, M. Explanation and Scientific Understanding. **Journal of Philosophy**, v. 71, n. 1, p. 5-19, 1974. DOI:10.2307/2024924.

FRIGG, R. Fiction in Science. *In:* WOODS, J. (ed.). **Fictions and Models**: New Essays. Munich: Philosophia Verlag, 2010a, p. 247-287.

FRIGG, R. Models and Fiction. **Synthese**, v. 172, n. 2, p. 251-268, 2010b. DOI:10.1007/s11229-009-9505-0.

FRIGG, R.; BRADLEY, S.; DU, H.; SMITH, L. A. Laplace's Demon and the Adventures of His Apprentices. **Philosophy of Science**, v. 81, n. 1, p. 31-59, 2014. DOI:10.1086/674416.

FRIGG, R.; NGUYEN, J. The Fiction View of Models Reloaded. **The Monist**, v. 99, n. 3, p. 225-242, 2016. DOI:10.1093/monist/onw002.

FRIGG, R.; NGUYEN, J. Mirrors without Warnings. **Synthese,** 2019. DOI:10.1007/s11229-019-02222-9.

FUMAGALLI, R. Why We Cannot Learn from Minimal Models. **Erkenntnis**, v. 81, n. 3, p. 433-455, 2016. DOI:10.1007/s10670-015-9749-7.

GÄHDE, U. Anomalies and the Revision of Theory-Elements: Notes on the Advance of Mercury's Perihelion. *In:* CHIARA, M. L. D.; DOETS, K.; MUNDICI, D.; VAN BENTHEM, J. (ed.). **Structures and Norms in Science** (Synthese Library 260). Dordrecht: Springer Netherlands, 1997, p. 89-104. DOI:10.1007/978-94-017-0538-7_6.

GALISON, P. **Image and Logic**: A Material Culture of Microphysics. Chicago: University of Chicago Press, 1997.

GELFERT, A. **How to Do Science with Models**: A Philosophical Primer (Springer Briefs in Philosophy). Cham: Springer International Publishing, 2016. DOI:10.1007/978-3-319-27954-1.

GENDLER, T. S. **Thought Experiment**: On the Powers and Limits of Imaginary Cases. New York: Garland, 2000.

GIBBARD, A.; VARIAN, H. R. Economic Models. **The Journal of Philosophy**, v. 75, n. 11, p. 664-677, 1978. DOI:10.5840/jphil1978751111.

GIERE, R. N. **Explaining Science**: A Cognitive Approach. Chicago: University of Chicago Press, 1988.

GIERE, R. N. **Science Without Laws.** Chicago: University of Chicago Press, 1999.

GIERE, R. N. **Scientific Perspectivism**. Chicago: University of Chicago Press, 2006.

GIERE, R. N. Why Scientific Models Should Not be Regarded as Works of Fiction. *In:* SUÁREZ, M. (ed.). **Fictions in Science**: Philosophical Essays on Modeling and Idealization. London: Routledge, 2009, p. 248-258.

DOI:10.4324/9780203890103.

GIERE, R. N. An Agent-Based Conception of Models and Scientific Representation. **Synthese**, v. 172, n. 2, p. 269-281, 2010. DOI:10.1007/s11229-009-9506-z.

GODFREY-SMITH, P. The Strategy of Model-Based Science. **Biology & Philosophy**, v. 21, n. 5, p. 725-740, 2007. DOI:10.1007/s10539-006-9054-6.

GODFREY-SMITH, P. Abstractions, Idealizations, and Evolutionary Biology. *In:* BARBEROUSSE, A.; MORANGE, M.; PRADEU, T. (ed.). **Mapping the Future of Biology: Evolving Concepts and Theories** (Boston Studies in the Philosophy of Science 266). Dordrecht: Springer Netherlands, 2009, p. 47-56. DOI:10.1007/978-1-4020-9636-5_4.

GROENEWOLD, H. J. The Model in Physics. *In:* FREUDENTHAL, H. (ed.). **The Concept and the Role of the Model in Mathematics and Natural and Social Sciences**. Dordrecht: Reidel, 1961, p. 98-103. DOI:10.1007/978-94-010-3667-2_9.

GRÜNE-YANOFF, T. Learning from Minimal Economic Models. **Erkenntnis**, v. 70, n. 1, p. 81-99, 2009. DOI:10.1007/s10670-008-9138-6.

HACKING, I. **Representing and Intervening**: Introductory Topics in the Philosophy of Natural Science. Cambridge: Cambridge University Press, 1983. DOI:10.1017/CBO9780511814563.

HALE, S. C. Spacetime and the Abstract/Concrete Distinction. **Philosophical Studies**, v. 53, n.1, p. 85-102, 1988. DOI:10.1007/BF00355677.

HARRÉ, R. **Modeling**: Gateway to the Unknown (Studies in Multidisciplinarity 1). Amsterdam: Elsevier, 2004.

HARRIS, T. Data Models and the Acquisition and Manipulation of Data. **Philosophy of Science**, v. 70, n. 5, p. 1508-1517, 2003. DOI:10.1086/377426.

HARTMANN, S. Models as a Tool for Theory Construction: Some Strategies of Preliminary Physics. *In:* HERFEL, W.; KRAJEWSKI, W.; NIINILUOTO, I.; WOJCICKI, R. (ed.). **Theories and Models in Scientific Process** (Poznań Studies in the Philosophy of Science and the Humanities 44). Amsterdam: Rodopi, 1995, p. 49-67.

HARTMANN, S. The World as a Process: Simulations in the Natural and Social Sciences. *In:* HEGSELMANN, R.; MUELLER, U.; TROITZSCH, K. G. (ed.). **Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View** (Theory and Decision Library 23). Dordrecht: Springer Netherlands, 1996, p. 77-100. DOI:10.1007/978-94-015-8686-3_5.

HARTMANN, S. Idealization in Quantum Field Theory. *In:* SHANKS, N. (ed.). **Idealization in Contemporary Physics**. Amsterdam: Rodopi, 1998, p. 99-122.

HARTMANN, S. Models and Stories in Hadron Physics. *In:* MORGAN, M. S.; MORRISON, M. (ed.). **Models as Mediators**: Perspectives on Natural and Social Science. Cambridge: Cambridge University Press, 1999, p. 326-346. DOI:10.1017/CBO9780511660108.012.

HARTMANN, S. Effective Field Theories, Reductionism and Scientific Explanation. **Studies in History and Philosophy of Science Part B**: Studies in History and Philosophy of Modern Physics, v. 32, n. 2, p. 267-304, 2001. DOI:10.1016/S1355-2198(01)00005-3.

HARTMANN, S.; HOEFER, C.; BOVENS, L. (ed.). **Nancy Cartwright's Philosophy of Science** (Routledge Studies in the Philosophy of Science). New York: Routledge, 2008.

HEGSELMANN, R.; MUELLER, U.; TROITZSCH, K. G. (ed.). **Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View** (Theory and Decision Library 23). Dordrecht: Springer Netherlands, 1996. DOI:10.1007/978-94-015-8686-3.

HELMAN, D. H. (ed.). **Analogical Reasoning**: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy (Synthese Library 197). Dordrecht: Springer Netherlands, 1988. DOI:10.1007/978-94-015-7811-0.

HEMPEL, C. G. **Aspects of Scientific Explanation and Other Essays in the Philosophy of Science**. New York: Free Press, 1965.

HERFEL, W.; KRAJEWSKI, W.; NIINILUOTO, I.; WOJCICKI, R. (ed.). **Theories and Models in Scientific Process** (Poznań Studies in the Philosophy of Science and the Humanities 44). Amsterdam: Rodopi, 1995.

HESSE, M. **Models and Analogies in Science.** London: Sheed and Ward, 1963.

HESSE, M. Models and Analogy in Science. *In:* EDWARDS, P. (ed.). **Encyclopedia of Philosophy.** New York: Macmillan, 1967, p. 354-359.

HESSE, M. **The Structure of Scientific Inference.** London: Macmillan, 1974.

HODGES, W. **A Shorter Model Theory.** Cambridge: Cambridge University Press, 1997.

HOLYOAK, K.; THAGARD, P. **Mental Leaps**: Analogy in Creative Thought. Cambridge: MIT Press, 1995.

HOROWITZ, T.; MASSEY, G. J. (ed.). **Thought Experiments in Science and Philosophy**. Lanham: Rowman & Littlefield, 1991.

ISAAC, A. M. C. Modeling without Representation. **Synthese**, v. 190, n. 16, p. 3611-3623, 2013. DOI:10.1007/s11229-012-0213-9.

JEBEILE, J.; KENNEDY, A. G. Explaining with Models: The Role of Idealizations. **International Studies in the Philosophy of Science**, v. 29, n. 4, p. 383-392, 2015. DOI:10.1080/02698595.2015.1195143.

JONES, M. R. Idealization and Abstraction: A Framework. *In:* JONES, M. R.; CARTWRIGHT, N. (ed.). **Idealization XII**: Correcting the Model (Poznań Studies in the Philosophy of the Sciences and the Humanities 86). Amsterdam: Rodopi, 2005, p. 173-217. DOI:10.1163/9789401202732_010.

JONES, M. R.; CARTWRIGHT, N. (ed.). **Idealization XII**: Correcting the Model (Poznań Studies in the Philosophy of the Sciences and the Humanities 86). Amsterdam: Rodopi, 2005. DOI:10.1163/9789401202732.

KELVIN, W. T. B. **Notes of lectures on molecular dynamics and the wave theory of light**. Cambridge: MIT Press, 1987.

KHALIFA, K. **Understanding, Explanation, and Scientific Knowledge**. Cambridge: Cambridge University Press, 2017. DOI:10.1017/9781108164276.

KLEIN, D.; MARX, J.; FISCHBACH, K. Agent-Based Modeling in Social Science History and Philosophy: An Introduction. **Historical Social Research**, v. 43, n. 1 , p. 243-258, 2018.

KNUUTTILA, T. Models, Representation, and Mediation. **Philosophy of Science**, v. 72, n. 5, p. 1260-1271, 2005. DOI:10.1086/508124.

KNUUTTILA, T. Modelling and Representing: An Artefactual Approach to Model-Based Representation. **Studies in History and Philosophy of Science Part A**, v. 42, n. 2, p. 262-271, 2011. DOI:10.1016/j.shpsa.2010.11.034.

KROES, P. Structural Analogies Between Physical Systems. **The British Journal for the Philosophy of Science**, v. 40, n. 2, p. 145-154, 1989. DOI:10.1093/bjps/40.2.145.

LANGE, M. On 'Minimal Model Explanations: A Reply to Batterman and Rice. **Philosophy of Science**, v. 82, n. 2, p. 292-305, 2015. DOI:10.1086/680488.

LAVIS, D. A. Boltzmann, Gibbs, and the Concept of Equilibrium. **Philosophy of Science**, v. 75, n. 5, p. 682-692, 2008. DOI:10.1086/594514.

LAYMON, R. Scientific Realism and the Hierarchical Counterfactual Path from Data to Theory. **Proceedings of the Biennial Meeting of the Philosophy of Science Association**, v. 1, p. 107–121, 1982. DOI:10.1086/psaprocbienmeetp.1982.1.192660.

LAYMON, R. Idealizations and the Testing of Theories by Experimentation. *In:* ACHINSTEIN, P.; HANNAWAY, O. (ed.). **Observation, Experiment, and Hypothesis in Modern Physical Science.** Cambridge: MIT Press, 1985, p. 147-173.

LAYMON, R. Thought Experiments by Stevin, Mach and Gouy: Thought Experiments as Ideal Limits and Semantic Domains. In: HOROWITZ, T.; MASSEY, G. J. (ed.). **Thought Experiments in Science and Philosophy**. Lanham: Rowman & Littlefield, 1991, p. 167-191.

LEONELLI, S. Packaging Small Facts for Re-Use: Databases in Model Organism Biology. *In:* HOWLETT, P.; MORGAN, M. S. (ed.). **How Well Do Facts Travel?** The Dissemination of Reliable Knowledge. Cambridge: Cambridge University Press, 2010, p. 325-348. DOI:10.1017/CBO9780511762154.017.

LEONELLI, S. **Data-Centric Biology**: A Philosophical Study. London: University of Chicago Press, 2016.

LEONELLI, S. What Distinguishes Data from Models? European Journal for **Philosophy of Science**, v. 9, n. 2, art 22, 2019. DOI:10.1007/s13194-018-0246-0.

LEONELLI, S.; ANKENY, R. A. Re-Thinking Organisms: The Impact of Databases on Model Organism Biology. **Studies in History and Philosophy of Science Part C**: Studies in History and Philosophy of Biological and Biomedical Sciences, v. 43, n. 1, p. 29-36, 2012. DOI:10.1016/j.shpsc.2011.10.003.

LEPLIN, J. The Role of Models in Theory Construction. *In:* NICKLES, T. (ed.). **Scientific Discovery, Logic, and Rationality** (Boston Studies in the Philosophy of Science 56). Dordrecht: Springer Netherlands, 1980, p. 267–283. DOI:10.1007/978-94-009-8986-3_12.

LEVY, A. Models, Fictions, and Realism: Two Packages. **Philosophy of Science**, v. 79, n. 5, p. 738-748, 2012. DOI:10.1086/667992.

LEVY, A. Modeling without Models. **Philosophical Studies**, v. 172, n. 3, p. 781-798, 2015. DOI:10.1007/s11098-014-0333-9.

LEVY, A.; CURRIE, A. Model Organisms Are Not (Theoretical) Models. **The British Journal for the Philosophy of Science**, v. 66, n. 2, p. 327-348, 2015. DOI:10.1093/bjps/axt055.

LEVY, A.; Peter GODFREY-SMITH (ed.). **The Scientific Imagination**: Philosophical and Psychological Perspectives. New York: Oxford University Press, 2020.

LIEFKE, K.; HARTMANN, S. Intertheoretic Reduction, Confirmation, and Montague's Syntax–Semantics Relation. **Journal of Logic, Language and Information**, v. 27, n. 4, p. 313-341, 2018. DOI:10.1007/s10849-018-9272-8.

LIPTON, P. Understanding without Explanation. *In:* DE REGT, H.; LEONELLI, S.; EIGNER, K. (ed.). **Scientific Understanding**: Philosophical Perspectives. Pittsburgh: University of Pittsburgh Press, 2009, p. 43-63.

LUCZAK, J.Talk about Toy Models. **Studies in History and Philosophy of Science** Part B: Studies in History and Philosophy of Modern Physics, v. 57, p. 1-7, 2017. DOI:10.1016/j.shpsb.2016.11.002.

MAGNANI, L. Scientific Models Are Not Fictions: Model-Based Science as Epistemic Warfare. *In:* MAGNANI, L.; LI, P. (ed.). **Philosophy and Cognitive Science**: Western & Eastern Studies (Studies in Applied Philosophy, Epistemology and Rational Ethics 2). Berlin: Springer, 2012, p. 1-38. DOI:10.1007/978-3-642-29928-5_1.

MAGNANI, L.; CASADIO, C. (ed.). **Model-Based Reasoning in Science and Technology**: Logical, Epistemological, and Cognitive Issues (Studies in Applied Philosophy, Epistemology and Rational Ethics 27). Cham: Springer International Publishing, 2016. DOI:10.1007/978-3-319-38983-7.

MAGNANI, L.; NERSESSIAN, N. J. (ed.). **Model-Based Reasoning**: Science, Technology, Values. Boston: Springer US, 2002. DOI:10.1007/978-1-4615-0605-8.

MAGNANI, L.; NERSESSIAN, N. J.; THAGARD, P. (ed.). **Model-Based Reasoning in Scientific Discovery.** Boston: Springer US, 1999. DOI:10.1007/978-1-4615-4813-3.

MÄKI, U. Isolation, Idealization and Truth in Economics. *In:* HAMMINGA, B.; MARCHI, N. B. D. (ed.). **Idealization VI**: Idealization in Economics (Poznań Studies in the Philosophy of the Sciences and the Humanities 38). Amsterdam: Rodopi, 1994, p. 147-168.

MASSIMI, M. Perspectivism. In: SAATSI, J. (ed.). The Routledge Handbook of Scientific Realism. London: Routledge, 2017, p. 164-175.

MASSIMI, M. Four Kinds of Perspectival Truth. **Philosophy and Phenomenological Research**, v. 96, n. 2, p. 342-359, 2018a. DOI:10.1111/phpr.12300.

MASSIMI, M. Perspectival Modeling. **Philosophy of Science**, v. 85, n. 3, p. 335-359, 2018b. DOI:10.1086/697745.

MASSIMI, M. Two Kinds of Exploratory Models. **Philosophy of Science**, v. 86, n. 5, p. 869-881, 2019. DOI:10.1086/705494.

MASSIMI, M.; MCCOY, C. D. (ed.). **Understanding Perspectivism**: Scientific Challenges and Methodological Prospects. New York: Routledge, 2019. DOI:10.4324/9781315145198.

MAYO, D. **Error and the Growth of Experimental Knowledge.** Chicago: University of Chicago Press, 1996.

MAYO, D. **Statistical Inference as Severe Testing**: How to Get Beyond the Statistics Wars. Cambridge: Cambridge University Press, 2018. DOI:10.1017/9781107286184.

MCMULLIN, E. What Do Physical Models Tell Us?. *In:* ROOTSELAAR, B. V.; STAAL, J. F. (ed.). **Logic, Methodology and Philosophy of Science III** (Studies in Logic and the Foundations of Mathematics 52). Amsterdam: North Holland, 1968, p. 385-396. DOI:10.1016/S0049-237X(08)71206-0.

MCMULLIN, E. Galilean Idealization. **Studies in History and Philosophy of Science Part A**, v. 16, n. 3, p. 247-273, 1985. DOI:10.1016/0039-3681(85)90003-2.

MORGAN, M. S. Learning from Models. *In:* MORGAN, M. S.; MORRISON, M. (ed.). **Models as Mediators**: Perspectives on Natural and Social Science. Cambridge: Cambridge University Press, 1999, p. 347-388. DOI:10.1017/CBO9780511660108.013.

MORGAN, M. S.; BOUMANS, M. J. Secrets Hidden by Two-Dimensionality: The Economy as a Hydraulic Machine. *In:* CHADAREVIAN, S. de.; HOPWOOD, N. (ed.). **Model**: The Third Dimension of Science. Stanford: Stanford University Press, 2004, p. 369-401.

MORGAN, M. S.; MORRISON, M. (ed.). **Models as Mediators**: Perspectives on Natural and Social Science. Cambridge: Cambridge University Press, 1999. DOI:10.1017/CBO9780511660108.

MORRISON, M. Models as Autonomous Agents. *In:* MORGAN, M. S.; MORRISON, M. (ed.). **Models as Mediators**: Perspectives on Natural and Social Science. Cambridge: Cambridge University Press, 1999, p. 38-65. DOI:10.1017/CBO9780511660108.004.

MORRISON, M. **Unifying Scientific Theories**: Physical Concepts and Mathematical Structures. Cambridge: Cambridge University Press, 2000. DOI:10.1017/CBO9780511527333.

MORRISON, M. Approximating the Real: The Role of Idealizations in Physical Theory. *In:* JONES, M. R.; CARTWRIGHT, N. (ed.). **Idealization XII**: Correcting the Model (Poznań Studies in the Philosophy of the Sciences and the Humanities 86). Amsterdam: Rodopi, 2005, p. 145-172. DOI:10.1163/9789401202732_009.

MORRISON, M. Understanding in Physics and Biology: From the Abstract to the Concrete. *In:* DE REGT, H.; LEONELLI, S.; EIGNER, K. (ed.). **Scientific Understanding**: Philosophical Perspectives. Pittsburgh: University of Pittsburgh Press, 2009, p. 123-145.

MORRISON, M. Emergent Physics and Micro-Ontology. **Philosophy of Science**, v. 79, n. 1, p. 141-166, 2012. DOI:10.1086/663240.

MUSGRAVE, A. Unreal Assumptions' in Economic Theory: The F-Twist Untwisted. **Kyklos**, 34, n. 3, p. 377-387, 1981. DOI:10.1111/j.1467-6435.1981.tb01195.x.

NAGEL, E. **The Structure of Science**: Problems in the Logic of Scientific Explanation. New York: Harcourt, Brace and World, 1961.

NERSESSIAN, N. J. Model-Based Reasoning in Conceptual Change. *In:* MAGNANI, L.; NERSESSIAN, N. J.; THAGARD, P. (ed.). **Model-Based Reasoning in Scientific Discovery.** Boston: Springer US, 1999, p. 5-22. DOI:10.1007/978-1-4615-4813-3_1.

NERSESSIAN, N. J. **Creating Scientific Concepts**. Cambridge: MIT Press, 2010.

NGUYEN, J. It's Not a Game: Accurate Representation with Toy Models. **The British Journal for the Philosophy of Science**, 2019. DOI:10.1093/bjps/axz010.

NGUYEN, J.; FRIGG, R. Mathematics Is Not the Only Language in the Book of Nature. **Synthese**, 2017. DOI:10.1007/s11229-017-1526-5.

NORTON, J. D. Thought Experiments in Einstein's Work. *In:* HOROWITZ, T.; MASSEY, G. J. (ed.). **Thought Experiments in Science and Philosophy**. Lanham: Rowman & Littlefield, 1991, p. 129-148.

NORTON, J. D. Causation as Folk Science. **Philosopher's**, v. 3, n. 4, p. 1-22, 2003. Disponível em: http://hdl.handle.net/2027/spo.3521354.0003.004. Acesso em: 10 out. 2021.

NORTON, J. D. Approximation and Idealization: Why the Difference Matters. **Philosophy of Science**, v. 79, n. 2, p. 207-232, 2012. DOI:10.1086/664746.

NOWAK, L. **The Structure of Idealization**: Towards a Systematic Interpretation of the Marxian Idea of Science. Dordrecht: D. Reidel, 1979.

PALACIOS, P. Phase Transitions: A Challenge for Intertheoretic Reduction? **Philosophy of Science**, v. 86, n. 4, p. 612-640, 2019. DOI:10.1086/704974.

PESCHARD, I. Making Sense of Modeling: Beyond Representation. **European Journal for Philosophy of Science**, v. 1, n. 3, p. 335-352, 2011. DOI:10.1007/s13194-011-0032-8.

PICCININI, G.; CRAVER, C. Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches. **Synthese**, v. 183, n. 3, p. 283-311, 2011. DOI:10.1007/s11229-011-9898-4.

PINCOCK, C. Mathematics and Scientific Representation. Oxford: Oxford University Press, 2012. DOI:10.1093/acprof:oso/9780199757107.001.0001.

PINCOCK, C. Concrete Scale Models, Essential Idealization and Causal Explanation. **British Journal for the Philosophy of Science**, 2020. DOI:.

PORTIDES, D. P. The Relation between Idealisation and Approximation in Scientific Model Construction. **Science & Education**, v. 16, n. 7-8, p. 699-724, 2007. DOI:10.1007/s11191-006-9001-6.

PORTIDES, D. P. How Scientific Models Differ from Works of Fiction. *In:* MAGNANI, L.; CASADIO, C. (ed.). **Model-Based Reasoning in Science and Technology** (Studies in Applied Philosophy, Epistemology and Rational Ethics 8). Berlin: Springer, 2014, p. 75-87. DOI:10.1007/978-3-642-37428-9_5.

POTOCHNIK, A. Optimality Modeling and Explanatory Generality. **Philosophy of Science**, v. 74, n. 5, p. 680-691, 2007.

POTOCHNIK, A. **Idealization and the Aims of Science.** Chicago: University of Chicago Press, 2017.

POZNIC, M. Make-Believe and Model-Based Representation in Science: The Epistemology of Frigg's and Toon's Fictionalist Views of Modeling. **Teorema: Revista Internacional de Filosofía**, v. 35, n. 3, p. 201-218, 2016.

PSILLOS, S. The Cognitive Interplay between Theories and Models: The Case of 19th Century Optics. *In:* HERFEL, W.; KRAJEWSKI, W.; NIINILUOTO, I.; WOJCICKI, R. (ed.). **Theories and Models in Scientific Process** (Poznań Studies in the Philosophy of Science and the Humanities 44). Amsterdam: Rodopi, 1995, p. 105-133.

REDHEAD, M. Models in Physics. **The British Journal for the Philosophy of Science**, v. 31, n. 2, p. 145-163, 1980. DOI:10.1093/bjps/31.2.145.

REISS, J. Causal Inference in the Abstract or Seven Myths about Thought Experiments. *In:* **Causality**: Metaphysics and Methods Research Project, Technical Report 03/02. London: London School of Economics, 2003.

REISS, J. Social Capacities. *In:* HARTMANN, S.; HOEFER, C.; BOVENS, L. (ed.). **Nancy Cartwright's Philosophy of Science** (Routledge Studies in the Philosophy of Science). New York: Routledge, 2006, p. 265-288.

REISS, J. The Explanation Paradox. **Journal of Economic Methodology**, v. 19, n. 1, p. 43-62, 2012. DOI:10.1080/1350178X.2012.661069.

REUTLINGER, A. Do Renormalization Group Explanations Conform to the Commonality Strategy? **Journal for General Philosophy of Science**, v. 48, n. 1, p. 143-150, 2017. DOI:10.1007/s10838-016-9339-7.

REUTLINGER, A.; HANGLEITER, D.; HARTMANN, S. Understanding (with) Toy Models. **The British Journal for the Philosophy of Science**, v. 69, n. 4,

p. 1069-1099, 2018. DOI:10.1093/bjps/axx005.

RICE, C. C. Moving Beyond Causes: Optimality Models and Scientific Explanation. **Noûs**, v. 49, n. 3, p. 589-615, 2015. DOI:10.1111/nous.12042.

RICE, C. C. Factive Scientific Understanding without Accurate Representation. **Biology & Philosophy**, v. 31, n. 1, p. 81-102, 2016. DOI:10.1007/s10539-015-9510-2.

RICE, C. C. Idealized Models, Holistic Distortions, and Universality. **Synthese**, v. 195, n. 6, p. 2795-2819, 2018. DOI:10.1007/s11229-017-1357-4.

RICE, C. C. Models Don't Decompose That Way: A Holistic View of Idealized Models. **The British Journal for the Philosophy of Science**, v. 70, n. 1, p. 179-208, 2019. DOI:10.1093/bjps/axx045.

ROSALER, J. Local Reduction in Physics. Studies in History and Philosophy of Science Part B: **Studies in History and Philosophy of Modern Physics**, v. 50, p. 54-69, 2015. DOI:10.1016/j.shpsb.2015.02.004.

RUEGER, A. Perspectival Models and Theory Unification. **The British Journal for the Philosophy of Science**, v. 56, n. 3, p. 579-594, 2005. DOI:10.1093/bjps/axi128.

RUEGER, A.; SHARP, D. Idealization and Stability: A Perspective from Nonlinear Dynamics. *In:* SHANKS, N. (ed.). **Idealization in Contemporary Physics**. Amsterdam: Rodopi, 1998, p. 201-216.

SAATSI, J. Models, Idealisations, and Realism. *In:* IPPOLITI, E.; STERPETTI, F.; NICKLES, T. (ed.). **Models and Inferences in Science** (Studies in Applied Philosophy, Epistemology and Rational Ethics 25). Cham: Springer International Publishing, 2016, p. 173-189. DOI:10.1007/978-3-319-28163-6_10.

SAATSI, J.; REUTLINGER, A. Taking Reductionism to the Limit: How to Rebut the Antireductionist Argument from Infinite Limits. **Philosophy of Science**, v. 85, n. 3, p. 455-482, 2018. DOI:10.1086/697735.

SALIS, F. The New Fiction View of Models. **The British Journal for the Philosophy of Science**, 2019. DOI:10.1093/bjps/axz015.

SALMON, Wesley C. **Scientific Explanation and the Causal Structure of the World.** Princeton: Princeton University Press, 1984.

SCHAFFNER, K. F. The Watson–Crick Model and Reductionism. **The British Journal for the Philosophy of Science**, v. 20, n. 4, p. 325-348, 1969. DOI:10.1093/bjps/20.4.325.

SCHEIBE, E. **Die Reduktion physikalischer Theorien**: Ein Beitrag zur Einheit der Physik, Teil I: Grundlagen und elementare Theorie. Berlin: Springer, 1997

.

SCHEIBE, E. **Die Reduktion physikalischer Theorien**: Ein Beitrag zur Einheit der Physik, Teil II: Inkommensurabilität und Grenzfallreduktion. Berlin: Springer, 1999.

SCHEIBE, E. **Between Rationalism and Empiricism**: Selected Papers in the Philosophy of Physics, Brigitte Falkenburg (ed.). New York: Springer, 2001. DOI:10.1007/978-1-4613-0183-7.

SHANKS, N. (ed.). **Idealization in Contemporary Physics**. Amsterdam: Rodopi, 1998.

SHECH, E. Idealizations, Essential Self-Adjointness, and Minimal Model Explanation in the Aharonov–Bohm Effect. **Synthese**, v. 195, n. 11, p. 4839-4863, 2018. DOI:10.1007/s11229-017-1428-6.

SISMONDO, S.; GISSIS, S. (ed.). Modeling and Simulation. **Special Issue of Science in Context**, v. 12, n. 2, 1999.

SORENSEN, Roy A. **Thought Experiments.** New York: Oxford University Press, 1992. DOI:10.1093/019512913X.001.0001.

SPECTOR, M. Models and Theories. **The British Journal for the Philosophy of Science**, v. 16, n. 62, p. 121-142, 1965. DOI:10.1093/bjps/XVI.62.121.

STALEY, K. W. **The Evidence for the Top Quark**: Objectivity and Bias in Collaborative Experimentation. Cambridge: Cambridge University Press, 2004.

STERRETT, S. G. Models of Machines and Models of Phenomena. **International Studies in the Philosophy of Science**, v. 20, n. 1, p. 69-80, 2006. DOI:10.1080/02698590600641024.

STERRETT, S. G. Scale Modeling. *In:* MICHELFELDER, D.; DOORN, N. (ed.). **Routledge Handbook of Philosophy of Engineering**, n. 32, 2019. Disponível em: http://philsci-archive.pitt.edu/16054/. Acesso em: 08 out. 2021.

STREVENS, M. The Causal and Unification Approaches to Explanation Unified —Causally. **Noûs**, v. 38, n. 1, p. 154-176, 2004. DOI:10.1111/j.1468-0068.2004.00466.x.

STREVENS, M. **Depth**: An Account of Scientific Explanation. Cambridge: Harvard University Press, 2008.

STREVENS, M. **Tychomancy**: Inferring Probability from Causal Structure, Cambridge: Harvard University Press, 2013.

SUÁREZ, M. Scientific Representation: Against Similarity and Isomorphism. **International Studies in the Philosophy of Science**, v. 17, n. 3, p. 225-244, 2003. DOI:10.1080/0269859032000169442.

SUÁREZ, M. An Inferential Conception of Scientific Representation. **Philosophy of Science**, v. 71, n. 5, p. 767-779, 2004. DOI:10.1086/421415.

SUÁREZ, M. (ed.). **Fictions in Science**: Philosophical Essays on Modeling and Idealization. London: Routledge, 2009. DOI:10.4324/9780203890103.

SUGDEN, R. Credible Worlds: The Status of Theoretical Models in Economics. **Journal of Economic Methodology**, v. 7, n. 1, p. 1-31, 2000. DOI:10.1080/135017800362220.

SULLIVAN, E.; KHALIFA, K. Idealizations and Understanding: Much Ado About Nothing? **Australasian Journal of Philosophy**,v. 97, n. 4, p. 673-689, 2019. DOI:10.1080/00048402.2018.1564337.

SUPPE, F. Theory Identity. *In:* NEWTON-SMITH, W. H. (ed.). **A Companion to the Philosophy of Science**. Oxford: Wiley-Blackwell, 2000, p. 525-527.

SUPPES, P. A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. **Synthese**, v. 12, n. 2-3, 1960, p. 287-301. DOI:10.1007/978-94-010-3667-2_16.

SUPPES, P. Models of Data. *In:* NAGEL, E.; SUPPES, P.; TARSKI, A. (ed.). **Logic, Methodology and Philosophy of Science**: Proceedings of the 1960 International Congress. Stanford: Stanford University Press, 1962, p. 252-261.

SUPPES, P. **Studies in the Methodology and Foundations of Science**: Selected Papers from 1951 to 1969. Dordrecht: Reidel, 1969.

SUPPES, P. Statistical Concepts in Philosophy of Science. **Synthese**, v. 154, n. 3, p. 485-496, 2007. DOI:10.1007/s11229-006-9122-0.

SWOYER, C. Structural Representation and Surrogative Reasoning. **Synthese**, v. 87, n. 3, p. 449-508, 1991. DOI:10.1007/BF00499820.

TABOR, M. **Chaos and Integrability in Nonlinear Dynamics**: An Introduction. New York: John Wiley, 1989.

TELLER, P. Twilight of the Perfect Model. **Erkenntnis**, v. 55, n. 3, p. 393-415, 2001. DOI:10.1023/A:1013349314515.

TELLER, P. Critical Study: Nancy Cartwright's The Dappled World: A Study of the Boundaries of Science. **Noûs**, v. 36, n. 4, p. 699-725, 2002. DOI:10.1111/1468-0068.t01-1-00408.

TELLER, P. Fictions, Fictionalization, and Truth in Science. *In:* SUÁREZ, M. (ed.). **Fictions in Science**: Philosophical Essays on Modeling and Idealization. London: Routledge, 2009, p. 235-247. DOI:10.4324/9780203890103.

TELLER, P. Referential and Perspectival Realism, Spontaneous Generations: A **Journal for the History and Philosophy of Science**, v. 9, n. 1, p. 151-164, 2018. DOI:10.4245/sponge.v9i1.26990.

TEŠIĆ, M. Confirmation and the Generalized Nagel–Schaffner Model of Reduction: A Bayesian Analysis. **Synthese**, v. 196, n. 3, p. 1097-1129. DOI:10.1007/s11229-017-1501-1.

THOMASSON, A. L. **Fiction and Metaphysics.** New York: Cambridge University Press, 1999. DOI:10.1017/CBO9780511527463.

THOMASSON, A. L. If Models Were Fictions, Then What Would They Be?. In: LEVY, A.; GODFREY-SMITH, P. (ed.). **The Scientific Imagination**: Philosophical and Psychological Perspectives. New York: Oxford University Press, 2020, p. 51-74.

THOMSON-JONES, M. Models and the Semantic View. **Philosophy of Science**, v. 73, n. 5, p. 524-535, 2006. DOI:10.1086/518322.

THOMSON-JONES, M. Realism about Missing Systems. *In:* LEVY, A.; GODFREY-SMITH, P. (ed.). **The Scientific Imagination**: Philosophical and Psychological Perspectives. New York: Oxford University Press, 2020, p. 75-101.

TOON, A. **Models as Make-Believe**: Imagination, Fiction and Scientific Representation. Basingstoke: Palgrave Macmillan, 2012.

TROUT, J. D. Scientific Explanation and the Sense of Understanding. **Philosophy of Science**, v. 69, n. 2, p. 212-233, 2002. DOI:10.1086/341050.

VAN FRAASSEN, B. C. **Laws and Symmetry**. Oxford: Oxford University Press, 1989. DOI:10.1093/0198248601.001.0001.

WALTON, K. L. **Mimesis as Make-Believe**: On the Foundations of the Representational Arts. Cambridge: Harvard University Press, 1990.

WEISBERG, M. Three Kinds of Idealization. **Journal of Philosophy**, v. 104, n. 12, p. 639-659, 2007. DOI:10.5840/jphil20071041240.

WEISBERG, M. **Simulation and Similarity**: Using Models to Understand the World. Oxford: Oxford University Press, 2013. DOI:10.1093/acprof:oso/9780199933662.001.0001.

WEISBERG, M.; MULDOON, R. Epistemic Landscapes and the Division of Cognitive Labor. **Philosophy of Science**, v. 76, n. 2, p. 225-252, 2009. DOI:10.1086/644786.

WIMSATT, W. False Models as Means to Truer Theories. *In:* NITECKI, M.; HOFFMAN, A. (ed.). **Neutral Models in Biology.** Oxford: Oxford University Press, 1987, p. 23-55.

WIMSATT, W. **Re-Engineering Philosophy for Limited Beings**: Piecewise Approximations to Reality. Cambridge: Harvard University Press, 2007.

WOODWARD, J. **Making Things Happen**: A Theory of Causal Explanation. Oxford: Oxford University Press, 2003. DOI:10.1093/0195155270.001.0001.

WOODY, A. I. More Telltale Signs: What Attention to Representation Reveals about Scientific Explanation. **Philosophy of Science**, v. 71, n. 5, p. 780-793, 2004. DOI:10.1086/421416.

ZOLLMAN, K. J. S. The Communication Structure of Epistemic Communities. **Philosophy of Science**, v. 74, n. 5, p.574-587, 2007. DOI:10.1086/525605.

(VII) Mecanismos na Ciência*

Autores: Carl Craver e James Tabery
Tradução: Renata M. S. Arruda
Revisão: Lucia C. Neco

Por volta da virada do século XXI, o que passou a ser chamado de nova filosofia mecanicista (ou, para abreviar, o novo mecanismo) surgiu como um marco estrutural para se pensar sobre as suposições filosóficas subjacentes a muitas áreas da ciência, especialmente na biologia, na neurociência e na psicologia. Nesse verbete, introduzimos e resumimos as características distintivas desse marco, e discutimos como ele aborda uma série de questões clássicas na filosofia da ciência, incluindo explicação, metafísica, as relações entre disciplinas científicas, e o processo de descoberta científica. Para cada uma dessas questões, mostramos como a concepção mecanística reorientou o trabalho filosófico, como o novo mecanismo contribuiu para a discussão e o que ainda resta a ser feito.

---

* CRAVER, C; TABERY, J. Mechanisms in Science. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Summer Edition. Stanford, CA: The Metaphysics Research Lab, 2019. Disponível em: https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/. Acesso em: 22 set. 2021.

The following is the translation of the entry on Mechanisms in Science by Carl Craver and James Tabery, 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/sum2019/entries/science-mechanisms/. 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/science-mechanisms/. 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. A ascensão do novo mecanismo

A filosofia da ciência do século XX foi amplamente dominada pelo empirismo lógico. Oferecendo mais uma estrutura para fazer filosofia da ciência do que um conjunto coerente de doutrinas, o empirismo lógico abordou uma série de questões na filosofia da ciência através das lentes das estruturas lógicas e matemáticas constitutivas do pensamento e da prática científica. O empirismo lógico tendia, em ampla medida, a focar em aspectos abstratos e epistêmicos da ciência, com pouca atenção à prática científica. A física foi o modelo dominante.

A nova filosofia mecânica surgiu na virada do século XXI como um novo marco para pensar sobre a filosofia da ciência. Os filósofos que desenvolveram essa concepção foram, em comparação com os empiristas lógicos, também profissionais da história da ciência, e focaram, em ampla medida, nas ciências biológicas, ao invés das físicas. Muitos novos mecanicistas desenvolveram seus sistemas explicitamente como sucessores das abordagens lógico-empiristas acerca de causalidade, níveis, explicação, leis da natureza, redução e descoberta.

Como no empirismo lógico, a nova filosofia mecanicista é antes uma orientação à temática da filosofia da ciência do que um conjunto sistemático e coerente de doutrinas. A abordagem surgiu quando filósofos e historiadores da ciência começaram a romper com a prática, outrora padrão, de reconstruir inferências científicas com as ferramentas da lógica e, em vez disso, adotaram a investigação detalhada de episódios reais da história da ciência. As principais doutrinas do empirismo lógico estavam sob intensa crítica há décadas, e uma nova era da filosofia da ciência atenta à dimensão histórica ganhou atenção através dos trabalhos de, por exemplo, Kuhn (1962), Laudan (1977), e Lakatos (1977). Para muitos desses estudiosos formados em um ambiente empirista pós-lógico, parecia que grande parte da prática da ciência contemporânea (tanto no laboratório quanto no papel) era impulsionada pela busca de mecanismos, que muitas das grandes conquistas na história da ciência foram descobertas de mecanismos, e que a filosofia mais tradicional da ciência, por algum motivo, havia fracassado em avaliar esse aspecto central da visão de mundo científica.

Alguns aspectos da nova filosofia mecanicista começaram a surgir no final da década de 1960. Fodor (1968), por exemplo, contrastou explicações mecanísticas (tratando de partes e suas interações com caráter de lei) com explicações funcionais em psicologia. Wimsatt (1972a, 1976), com base no trabalho de Simon (1962) e Kaufman (1971), argumentou repetidamente que as estruturas abstratas e idealizadas

do empirismo lógico eram pouco adequadas para entender como os cientistas descobrem e explicam sistemas complexos em múltiplos níveis de organização. Cummins (1975) forneceu uma explicação de análises funcionais, caracterizando uma função como uma contribuição que uma parte de um componente faz para a capacidade geral de algum sistema que inclui esse componente. Salmon (1984, 1989) argumentou que as concepções empiricistas da explicação científica são fundamentalmente falhas porque negligenciam mecanismos causais. Cartwright (1989) argumentou que a concepção lógica empiricista de uma lei da natureza é, de fato, uma ficção filosófica usada para descrever a busca por capacidades e máquinas nomológicas.

Essas vertentes começaram a se unir em uma perspectiva abrangente nos anos 90. A primeira declaração clara do novo mecanismo foi em *Discovering Complexity* [Descobrindo Complexidade] de Bechtel e Richardson (2010). Eles conscientemente deixaram de lado as preocupações lógico-empiricistas com redução de teorias e passaram a se concentrar no processo pelo qual os cientistas descobrem mecanismos (consulte a Seção 6 abaixo). Logo depois, Glennan argumentou que mecanismos são a conexão secreta que Hume buscou entre causa e efeito (1996), uma tese relacionada ao, e parcialmente inspirada pelo, foco de Cartwright em capacidades e máquinas nomológicas (GLENNAN, 1997). Da mesma forma, em *How Scientists Explain Diseases* [Como cientistas explicam as doenças], Thagard centrou a busca por causas e mecanismos na medicina (THAGARD, 2000; *vide* a Seção 6 abaixo). O artigo "*Thinking about Mechanisms*" (Pensando sobre Mecanismos) de Machamer, Darden e Craver (Machamer, Darden e Craver 2000; familiarmente conhecido como "MDC") reuniu essas vertentes e se tornou para muitos o foco de atenção da nova perspectiva mecanística. A sugestão dos filósofos MDC foi a de que a filosofia da biologia, e talvez a filosofia da ciência em geral, devesse ser reestruturada em torno da ideia fundamental de que muitos cientistas organizam seu trabalho em busca de mecanismos.

2. O conceito de mecanismo

O termo "mecanismo" surgiu no século XVII e se derivou dos termos gregos e latinos para "máquina" (DIJKSTERHUIS, 1961). Descartes entendeu a mecânica como o componente fundamental do mundo físico; em *Le Monde*, ele propôs explicar diversos fenômenos no mundo natural (como o movimento planetário, as marés, o

movimento do sangue e as propriedades da luz) em termos de conservação do movimento inercial por meio do contato. Posteriormente, a ideia de mecanismo foi muitas vezes transformada para refletir uma evolução da compreensão acerca das forças causais básicas no mundo (além do movimento conservado): por exemplo, atração e repulsão (du Bois Reymond), conservação de energia (Helmholtz), atração gravitacional (Newton) (BOAS, 1952; WESTFALL, 1971; *vide* verbetes sobre Hermann von Helmholtz e Isaac Newton). O conceito de mecanismo tem tido uma evolução quase separada na história das ciências da vida (ALLEN, 2005; DES CHENE, 2001, 2005; NICHOLSON, 2012), às vezes evitando a austeridade metafísica adotada por Descartes e muitos dos primeiros mecanicistas.

Os novos mecanicistas herdam a palavra "mecanismo" desses antecedentes, mas, em seu esforço para capturar como o termo é usado na ciência contemporânea, distanciaram-se tanto da ideia de que mecanismos são máquinas, e, especialmente, da austera representação metafísica do mundo na qual toda mudança real envolve apenas uma, ou um conjunto limitado de, atividades ou forças fundamentais (ANDERSEN, 2014a, b).

Os mecanicistas geralmente evitam o esforço de definir as condições necessárias e suficientes para que algo seja um mecanismo. Em vez disso, eles oferecem descrições qualitativas construídas para capturar a maneira como os cientistas usam o termo e para implementar o conceito em suas práticas experimentais e inferenciais.

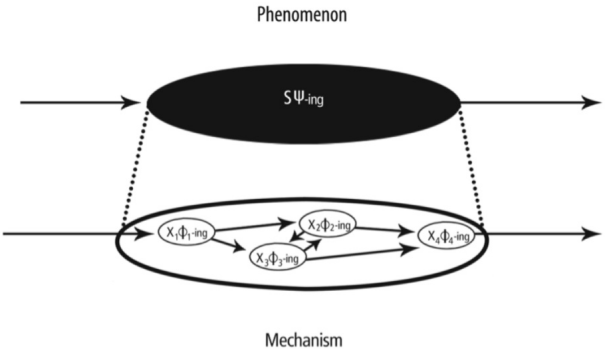
Três caracterizações são as mais comumente citadas:

> ● MDC: "Mecanismos são entidades e atividades organizadas de forma tal que produzem mudanças regulares a partir de condições de início ou de instalação até condições de fim ou término" (2000, p. 3).
> ● Glennan: "Um mecanismo para um comportamento é um sistema complexo que produz esse comportamento pela interação de várias partes, onde a interação entre as partes pode ser caracterizada por generalizações diretas, invariantes e relacionadas à mudança" (2002, p. S344).
> ● Bechtel e Abrahamsen: "Um mecanismo é uma estrutura que desempenha uma função em virtude de suas partes componentes, de suas operações componentes e de sua organização. O funcionamento

> orquestrado do mecanismo é responsável por um ou mais fenômenos" (2005, p. 423).

Cada uma dessas caracterizações contém quatro características básicas: (1) um fenômeno, (2) partes, (3) causas e (4) organização. Consideramos cada uma delas em mais detalhes abaixo.

Uma representação visual canônica de um mecanismo subjacente a um fenômeno é mostrada na Figura 1 (DE CRAVER, 2007). No topo está o fenômeno, algum sistema *S* envolvido no comportamento ψ. Esse é o comportamento do mecanismo como um todo. Abaixo estão as partes (os X_s) e suas atividades (os φ_s) organizadas juntas. As linhas pontilhadas quase na vertical refletem o fato de que as partes e atividades estão contidas e são componentes do mecanismo envolvido nesse comportamento. Assim representado, os mecanismos são decomposicionais no sentido de que o comportamento do sistema como um todo pode ser dividido em interações organizadas entre as atividades das partes.



Nos primeiros trabalhos deste tópico, essas diferentes caracterizações eram frequentemente tratadas como concorrentes. Tabery (2004) argumentou que, ao contrário, elas refletem ênfases e orientações intelectuais diferentes e complementares. Muitos mecanicistas adotaram essa postura ecumênica. Por exemplo, Illari e Williamson oferecem um "conceito consensual" de mecanismo:

> Um mecanismo para um fenômeno consiste em entidades e atividades organizadas de maneira tal

> que elas sejam responsáveis pelo fenômeno (2012, p. 120)

Da mesma forma, Glennan se refere ao "mecanismo mínimo":

> Um mecanismo para um fenômeno consiste em entidades (ou partes) cujas atividades e interações são organizadas de maneira tal que elas produzam o fenômeno (GLENNAN, cap. 2, no prelo).

Cada uma dessas caracterizações incluem os quatro elementos básicos e são projetadas para tornar a caracterização mais inclusiva. A insistência de MDC na regularidade dos mecanismos é abandonada, por exemplo, para acomodar mecanismos que funcionam apenas uma vez ou que funcionam irregularmente (SKIPPER; MILSTEIN, 2005; BOGEN, 2005; *vide* Seção 2.1.2 abaixo). A ênfase de Bechtel e Abrahamsen nas "funções" é abandonada para acomodar mecanismos que não servem a nenhum fim e para distanciar o mecanismo desse termo carregado de significados, tão frequentemente oposto ao mecanismo (CRAVER, 2001a; GARSON, 2013; MALEY; PICCININI, no prelo; *vide* Seção 4.5 abaixo).

Essas caracterizações ecumênicas minimizam intencionalmente o fato de que o termo "mecanismo" é usado de maneira diferente em diferentes contextos científicos e filosóficos. Confira Levy (2013) e, para visões alternativas das diferenças, veja Andersen (2014a, 2014b). De fato, grande parte do progresso nos anos anteriores envolveu aprender a reconhecer as muitas maneiras pelas quais o termo "mecanismo" pode ser usado e os muitos compromissos que podem ser assumidos em seu nome. Para ainda outras caracterizações de mecanismo, consulte Woodward (2002), Fagan (2012), Nicholson (2012) e Garson (2013). Tomando essas concepções ecumênicas como ponto de partida, vamos considerar os quatro componentes básicos: 1) o fenômeno, 2) partes, 3) causas e 4) organização.

2.1 Fenômeno

O fenômeno é o comportamento do mecanismo como um todo. Todos os mecanismos são mecanismos de algum fenômeno (KAUFFMAN, 1971; GLENNAN, 1996, 2002). O mecanismo da síntese de proteínas sintetiza proteínas. O mecanismo

do potencial de ação gera potenciais de ação. Os limites de um mecanismo – o que está no mecanismo e o que não está – são fixados em referência ao fenômeno que o mecanismo explica. Os componentes de um mecanismo são componentes em virtude de serem relevantes para o fenômeno.

MDC (2000) descrevem o funcionamento de mecanismos a partir de condições de início ou instalação até condições de término. Eles insistem que a descrição do fenômeno como uma relação entre *input* e *output* é rasa, porque muitas vezes existem vários inputs e outputs em um mecanismo, e porque as características centrais de um fenômeno podem não ser nem inputs nem outputs (mas sim detalhes sobre como o fenômeno se desdobra ao longo do tempo). Darden, apelando para o exemplo da síntese de proteínas, frequentemente associa o fenômeno ao estado final: a proteína (Darden 2006). Craver (2007), seguindo Cummins (1975) e Cartwright (1989), falam frequentemente do fenômeno como, grosso modo, uma capacidade ou comportamento do mecanismo como um todo.

2.1.1 Produzir, Subjazer e Manter

Os novos mecanicistas falam de diferentes maneiras do mecanismo como produzindo, subjazendo ou mantendo o fenômeno (CRAVER; DARDEN, 2013). A linguagem de **produção** é mais bem aplicada a mecanismos concebidos como uma sequência causal que termina em algum produto final: como quando um vírus produz sintomas através de um mecanismo de doença ou uma enzima fosforila um substrato. Em tais casos, o fenômeno pode ser um objeto (a produção de uma proteína), um estado de coisas (ser fosforilado), ou uma atividade ou evento (como a digestão). Para muitos mecanismos fisiológicos, por outro lado, é mais apropriado dizer que o mecanismo **subjaz** o fenômeno. O mecanismo do potencial de ação ou da memória de trabalho, por exemplo, subjaz o fenômeno, aqui caracteristicamente entendido como uma capacidade ou comportamento do mecanismo como um todo. Finalmente, um mecanismo pode **manter** um fenômeno, como quando os mecanismos homeostáticos mantêm a temperatura do corpo dentro de limites bem circunscritos. Nesses casos, o fenômeno é um estado de coisas, ou talvez um intervalo de estados de coisas, mantido pelo mecanismo. Essas expressões podem em muitos casos ser intertraduzidas (por exemplo, o produto é produzido, a produção tem um mecanismo subjacente, e o estado das coisas é mantido por um mecanismo subjacente). No entanto, claramente pode surgir uma confusão ao se misturar essas expressões.

2.1.2 Regularidade

A relação entre o mecanismo e o fenômeno deve ser regular? Esta é uma área de discussão ativa (DESAUTELS, 2011; ANDERSEN, 2011, 2014a, 2014b; KRICKEL, 2014). MDC estipulam que os mecanismos são regulares, posto que eles trabalham "sempre ou na maior parte do mesmo modo, nas mesmas condições" (2000, p. 3). Alguns entenderam isso (incorretamente, em nossa opinião) como afirmando que não existem mecanismos que funcionam apenas uma vez, ou que um mecanismo deve funcionar significativamente mais de uma vez para contar como um mecanismo.

Alguns argumentam que os mecanismos devem ser regulares nesse sentido factual (ANDERSEN, 2014a, 2014b); isto é, repetido em várias ocasiões (LEURIDAN, 2010). Essa concepção parece exigir um ponto de corte um tanto arbitrário no grau de regularidade entre coisas que verdadeiramente contam como mecanismos e aquelas que não contam. Alguns mecanicistas (BOGEN, 2005; GLENNAN, 2009) argumentam que não há dificuldade em aplicar o termo "mecanismo" a sequências causais pontuais, como quando um historiador fala do mecanismo que deu origem à Primeira Guerra Mundial. Outros mecanicistas argumentam que a distinção tipo-instância/espécime [*type-token*] é uma dicotomia muito grosseira para capturar os muitos níveis de abstração com os quais os tipos e as instâncias/espécimes do mecanismo podem ser caracterizados (DARDEN, 1991).

É possível, no entanto, ler a afirmação de MDC como asseverando não um tipo de regularidade factual, mas como um tipo contrafactual de quase-determinismo: fossem todas as condições iguais, então o mecanismo provavelmente produziria o mesmo fenômeno, no qual o termo "provavelmente" acomoda mecanismos com elementos estocásticos.

Embora a abordagem de MDC deixe aberta a possibilidade de que alguns mecanismos sejam estocásticos, ela descarta claramente mecanismos que geralmente não chegam a produzir seus fenômenos. Skipper e Millstein (2005) reforçam esse ponto para argumentar que a abordagem de MDC não pode acolher a ideia de que a seleção natural é um mecanismo. Se, como Gould (1990) argumentou, não se pudesse assistir a história da vida rebobinando as fitas e deixando tudo avançar novamente, então a seleção natural não seria um mecanismo MDC (*vide* Seção 2.6 abaixo). Não está claro por que MDC permitiriam a possibilidade de mecanismos

estocásticos e excluiriam, por definição, a possibilidade de que eles possam falhar com mais frequência do que funcionam. Se algum mecanismo biológico é verdadeiramente irregular nesse sentido (isto é, todos os fatores causalmente relevantes são os mesmos, mas o produto do mecanismo difere) é uma questão distinta de se eles são mecanismos em sentido absoluto (BOGEN, 2005; MACHAMER, 2004). Já Steel (2008) desenvolve uma abordagem estocástica dos mecanismos.

Krickel (2014) revisa as muitas maneiras diferentes de desempacotar a noção relevante de regularidade (*vide* ANDERSEN, 2012). Sua solução preferida, "regularidade reversa", sustenta que deve haver uma generalização no sentido de que, usualmente, quando o fenômeno ocorre, o mecanismo estava agindo.

2.2 Partes

Os mecanicistas têm se esforçado para encontrar uma maneira concisa de expressar a ideia de relação entre as partes [*parthood*] exigida dos componentes em um mecanismo. O plano é desenvolver uma abordagem que seja suficientemente permissiva para incluir os mecanismos paradigmáticos de diversas áreas da ciência, porém que não seja vazia.

Mereologias formais são difíceis de se aplicarem às partes materiais dos mecanismos biológicos. Axiomas da mereologia, como reflexividade (tudo faz parte de si mesmo) e composição irrestrita (quaisquer duas coisas formam um todo) não se aplicam aos padrões biológicos de uso do conceito de "parte".

Glennan (1996) reconheceu a dificuldade de definir a relação entre as partes logo de início. Sua proposta:

> As partes dos mecanismos devem ter um tipo de robustez e realidade, além de seu lugar dentro desse mecanismo. Deveria ser possível, em princípio, retirar a parte do mecanismo e considerar suas propriedades em outro contexto (1996, p. 53).

No entanto, mesmo isso seja talvez forte demais, já que algumas partes de um mecanismo podem se tornar instáveis quando removidas de seu contexto mecanístico. Posteriormente, Glennan (2002, p. S345) afirma que as propriedades de uma parte devem ser estáveis na ausência de intervenções, ou que as partes

devem ser estáveis o suficiente para serem chamadas de objetos. Essa noção talvez seja forte demais para acomodar as partes mais efêmeras de alguns mecanismos bioquímicos ou dos mecanismos de seleção natural (SKIPPER; MILLSTEIN, 2005; ILLARI; WILLIAMSON, 2010).

2.3 Causas

Mecanicistas discordaram entre si sobre como entender a causa no mecanismo causal. Os novos mecanicistas geralmente têm se esforçado (1) para liberar a noção causal relevante de qualquer concepção excessivamente austera que restrinja a causalidade a apenas uma pequena classe de fenômenos (como colisões, atração/repulsão, ou conservação de energia) (2) e ainda para distanciar-se da concepção de causalidade humeana, regularista, comum entre os empiristas lógicos (ver também o verbete da metafísica da causalidade). Quatro maneiras de investigar a causa no mecanismo causal têm sido discutidas: as abordagens de quantidade conservadas, as abordagens mecanísticas, as abordagens de atividades e as abordagens contrafactuais (deve-se observar que alguns mecanicistas amadureceram suas reflexões sobre a causalidade).

2.3.1 Abordagens de Quantidade Conservada

De acordo as abordagens de transmissão, a causa envolve a transmissão e propagação de marcas ou quantidades conservadas (SALMON, 1984, 1994; DOWE, 1992). A forma mais influente dessa concepção sustenta que dois processos causais interagem causalmente quando se cruzam no espaço-tempo e trocam alguma porção de uma quantidade conservada, como a massa. Nessa concepção, a causalidade é local (os processos devem se interceptar) e singular (é completamente instanciada em processos causais particulares), embora a abordagem dependa de leis de conservação (HITCHCOCK, 1995). Ainda que essa concepção tenha inspirado muitos dos novos mecanicistas, e ainda que ela compartilhe o compromisso destes em procurar na ciência uma explicação da causalidade, ela tem sido geralmente rejeitada pelos novos mecanicistas (*vide* MILLSTEIN, 2006; ROE, 2014).

Essa concepção tem sido impopular em parte porque apresenta reduzida aplicação direta em ciências não fundamentais, como a biologia. As alegações

causais que os biólogos fazem geralmente não envolvem referência explícita a quantidades conservadas, mesmo que pressuponham fundamentalmente tais noções (GLENNAN, 2002; CRAVER, 2007). Além disso, os mecanismos biológicos frequentemente envolvem causalidade por omissão, prevenção, e dupla prevenção (ou seja, quando um mecanismo funciona removendo uma causa, prevenindo uma causa, ou inibindo um inibidor) (SCHAFFER, 2000, 2004). Tais formas de desconexão causal são onipresentes nas ciências especiais.

2.3.2 Abordagens Mecanísticas

Glennan (1996, 2009) entende a causalidade (pelo menos a causalidade não fundamental) como derivada do conceito de mecanismo: afirmações causais são afirmações sobre a existência de um mecanismo. O veritador ou produtor de verdade [*truth-maker*] para uma afirmação causal em um nível da organização é um mecanismo que se encontra em um nível inferior. Em resumo, mecanismos são a conexão oculta que Hume buscou entre causa e efeito. Tal como a abordagem de Salmon-Dowe, a concepção de Glennan é singular: mecanismos específicos vinculam causas e efeitos particulares (GLENNAN, no prelo).

Essa concepção tem sido acusada de circularidade: o conceito de mecanismo contém de forma não eliminável um elemento causal. No entanto, Glennan responde que muitas abordagens da causalidade (como a abordagem de Woodward em 2003, consulte a Seção 2.3.4) compartilham essa falha. Além disso, ele argumenta que, pelo menos para todas as causas não fundamentais, um mecanismo explica claramente como uma determinada causa produz seu efeito.

O êxito da análise depende de como se lida com a regressão resultante (CRAVER, 2007). Como observa Glennan (2009), a decomposição de causas em mecanismos pode continuar infinitamente, e assim não faz sentido argumentar sobre qual noção é mais fundamental; ou a decomposição pode se fundamentar em alguma noção causal básica, de nível mais baixo, que é primitiva e, portanto, não analisável em outros mecanismos causais. A última opção precisa desafiar a ausência de causalidade amplamente aclamada nas teorias da física fundamental (RUSSELL, 1913); em escalas de tamanho muito pequeno, as concepções clássicas de objetos e propriedades não parecem mais se aplicar, tornando difícil perceber que conteúdo resta à ideia de que existem mecanismos em funcionamento (*vide* TELLER, 2010; KUHLMAN; GLENNAN, 2014).

2.3.3 Abordagens Baseadas em Atividades

Outros mecanicistas, como Bogen (2005, 2008a) e Machamer (2004), adotam uma concepção não-redutora e anscombiana de que a causa deve ser entendida em termos de atividades produtivas. Atividades são tipos de causas, como atração e repulsão magnética ou ligações de hidrogênio. Os defensores de abordagens baseadas em atividades evitam a necessidade de definir o conceito, confiando na ciência para dizer o que são atividades e quais recursos elas podem ter. Essa concepção é uma espécie de minimalismo causal (GODFREY-SMITH, 2010). A ocorrência de uma causa não é uma questão de com que frequência ela ocorre ou se ocorreria sempre ou na maioria das vezes nas mesmas condições (BOGEN, 2005).

Essa abordagem foi criticada como vazia porque fracassa em dizer o que as atividades são (PSILLOS, 2004), em justificar a relação de relevância causal e explicativa (WOODWARD, 2002) e em marcar uma distinção adequada entre atividades e correlações (PSILLOS, 2004), porém confira Bogen (2005, 2008a) para uma resposta. Glennan (no prelo) argumenta que esses problemas podem ser tratados reconhecendo que as atividades de um mecanismo em um nível dependem de mecanismos de nível inferior. A esse respeito, confira, também, Persson (2010) para uma crítica de atividades baseada em sua incapacidade de lidar com casos de efeitos poligênicos.

2.3.4 Abordagens contrafactuais

Por fim, alguns novos mecanicistas, particularmente aqueles interessados em fornecer uma análise da explicação científica, se inclinaram em direção a uma concepção contrafactual de relevância causal e, em particular, à concepção manipulacionista expressa em Woodward (2001, 2003). Confira, também, Glennan (2002) e Craver (2007). O compromisso central dessa concepção é que modelos de mecanismos descrevem variáveis que fazem diferença nos valores de outras variáveis no modelo e para o fenômeno. Fazer diferença nesse sentido manipulacionista é entendido como uma relação entre variáveis nas quais intervenções nas variáveis de causa podem ser usadas para alterar o valor das variáveis de efeito.

Diferentemente das concepções discutidas acima, esse modo de pensar sobre causalidade fornece uma análise conveniente da relevância explicativa que é compatível com os métodos para testar afirmações causais. Grosso modo, uma variável é causalmente relevante para uma segunda quando existe uma intervenção ideal na primeira que altera o valor da segunda através da alteração induzida na primeira. A concepção acolhe facilmente omissões, prevenções e prevenções duplas – situações que tradicionalmente se mostraram problemáticas para abordagens do tipo produção de causalidade. Em suma, a alegação de que **C** causa **E** requer apenas que intervenções ideais em **C** possam ser usadas para alterar o valor de **E**, não que **C** e **E** estejam fisicamente conectados um ao outro. Finalmente, essa concepção fornece algumas ferramentas para acolher relações causais de nível superior e as leis não acidentais da biologia. Por outro lado, a abordagem contrafactual é não redutiva (como a concepção mecanística), e herda desafios enfrentados por outras visões contrafatuais, como preempção e superdeterminação, que são comuns em mecanismos biológicos.

2.4 Organização

A organização característica dos mecanismos é ela própria objeto de considerável discussão.

2.4.1 Organização e Agregatividade

Wimsatt (1997) contrasta organização mecanística e agregação, uma distinção que os mecanicistas usaram para articular como as partes de um mecanismo são organizadas para formar um todo (*vide* CRAVER, 2001b). Propriedades agregadas são propriedades de totalidades que são somas simples das propriedades de suas partes. Em agregados, as partes podem ser reorganizadas e intersubstituídas uma pela outra sem alterar a propriedade ou o comportamento do todo, o todo pode ser desmontado e voltar a ser reunido sem interromper a propriedade ou o comportamento do todo, e a propriedade do todo muda somente linearmente com a adição e remoção de partes. Essas características dos agregados são mantidas porque a organização é irrelevante para a propriedade do todo. Wimsatt, portanto, concebe a organização como não-agregatividade. Ele também a descreve como uma forma mecanística

de emergência (consulte a Seção 4.2).

A emergência mecanística é onipresente – propriedades verdadeiramente agregativas são raras. Assim, os mecanicistas tendem a reconhecer um espectro de organização, com agregados de um lado e mecanismos altamente organizados do outro. De fato, muitos mecanismos estudados por biólogos envolvem partes e causas em todo esse espectro. Para uma discussão mais aprofundada sobre a emergência mecanística em relação a outras variantes, consulte Richardson e Stephan (2007).

2.4.2 Variedades de organização

Seguindo Wimsatt, os mecanicistas detalharam vários tipos de organização, característicos dos mecanismos. Uma lista canônica inclui tanto organização espacial quanto temporal. A organização espacial inclui localização, tamanho, forma, posição e orientação; a organização temporal inclui a ordem, o ritmo e a duração das atividades do componente. Mais recentemente, os mecanicistas enfatizaram os padrões organizacionais nos mecanismos como um todo. Bechtel, por exemplo, discute como modelos matemáticos, e modelos dinâmicos em particular, são usados para revelar organização temporal complexa em mecanismos interativos (BECHTEL, 2006, 2011, 2013b). Alguns argumentam que os modelos dinâmicos extrapolam os limites da abordagem mecanística (CHEMERO; SILBESTEIN, 2008) e, às vezes, o próprio Bechtel (KAPLAN; BECHTEL, 2011). Outros argumentam que os modelos dinâmicos são, na verdade, meramente descritivos com frequência (ou seja, modelos não explicativos) ou, alternativamente, que são usados para descrever a organização temporal de mecanismos (KAPLAN; BECHTEL, 2011; KAPLAN, 2012).

Mecanicistas, recentemente, têm também se utilizado do trabalho de Alon (2006; MILO *et al*., 2002) sobre *network motifs* [redes funcionais], padrões repetidos em redes causais, para expandir o vocabulário para se pensar em padrões abstratos de organização (LEVY, 2014; LEVY; BECHTEL, 2012). Entender como as partes compõem totalidades provavelmente será uma área de crescimento no futuro da abordagem mecanística. Para algumas outras contribuições recentes, consulte Kuorikoski e Ylikoski (2013), Kuhlmann (2011) e Glennan (no prelo).

2.4.3 Modularidade

A definição contrafactual de Woodward (2001, 2002, 2011, 2014) de um mecanismo (que é indiretamente especificada por meio de uma abordagem de modelos mecanísticos), bem como a definição descendente elaborada por Menzies (2012), exigem que os modelos de mecanismos sejam modulares. Isso significa, grosso modo, que deveria ser fisicamente possível intervir em uma suposta variável de causa em um mecanismo sem perturbar as relações funcionais entre as outras variáveis no mecanismo. Em termos de modelos de equações estruturais em particular, isso significa que é possível substituir o lado direito de uma equação no modelo por um valor específico (isto é, definir a variável da esquerda com um valor) sem a necessidade de alterar quaisquer das outras equações no modelo. Com isso pretende-se capturar formalmente o sentido de o mecanismo ser composto de partes interativas e separáveis. Para argumentos a favor de uma condição de modularidade em modelos mecanísticos, consulte Menzies (2012).

Steel (2008) apela para uma forma um tanto mais fraca de modularidade em sua análise probabilística de mecanismos – forma que deriva diretamente da ideia de Simon (1996) de sistemas aproximadamente decomponíveis. Na visão de Simon, as partes de um mecanismo têm mais, e mais fortes, relações causais com outros componentes no mecanismo, do que com itens fora do mecanismo. Isso dá aos mecanismos (e as partes de mecanismos) um tipo de "independência" ou "objetização" (condição de objeto), definida em última análise em termos da intensidade da interação entre os componentes. Grush (2003), seguindo Haugeland (1998), desenvolve uma ideia de modularidade em termos da amplitude da taxa (ou banda) de interação, na qual os módulos têm uma alta taxa (banda) de interações internas e baixa taxa (banda) de interações externas. Nesta visão, a modularidade não é uma questão de tudo ou nada, mas uma questão de grau; mecanismos são apenas quase decompostos. Craver (2007) argumenta que essa noção genérica falha em justificar a relevância de diferentes interações causais para diferentes decomposições mecanísticas; o que conta como parte de um mecanismo só pode ser definido em relação a alguma decisão anterior sobre o que se considera que o mecanismo está fazendo. Para críticas à modularidade, consulte Mitchell (2005) e Cartwright (2001, 2002).

2.4.4 Junção

Fagan (2012, 2013) dá ênfase à relação interdependente entre partes de um mecanismo. Os componentes de um mecanismo, ela ressalta, geralmente formam uma unidade mais complexa em virtude das propriedades individuais que os unem – suas "propriedades de mesclagem"; a unidade complexa então aparece no comportamento do mecanismo. Essa relação interdependente – junção – é exemplificada pelo modelo de ação enzimática chave-e-fechadura. Fagan aplica essa noção à pesquisa em células-tronco (FAGAN, 2013), mas argumenta que é uma característica geral da biologia experimental (FAGAN, 2012).

2.4.5 Níveis

Muitos mecanicistas dão ênfase à organização hierárquica dos mecanismos e à estrutura multinível de teorias nas ciências especiais (CRAVER, 2007, cap. 5). Os antecedentes do novo mecanismo focaram quase que exclusivamente em relações etiológicas e causais. No entanto, a nova ênfase nos mecanismos na biologia e nas ciências especiais exigiu uma análise das relações mecanísticas entre diferentes níveis de organização.

De uma perspectiva mecanística, níveis não são divisórias monolíticas no mobiliário do universo (OPPENHEIM; PUTNAM, 1958), nem são eles fundamentalmente uma questão de tamanho ou a exclusividade de interações causais dentro de um nível (WIMSATT, 1976). Ao contrário, níveis de mecanismos são definidos localmente dentro de um mecanismo multinível: um item está em um nível mais baixo do que outro quando o primeiro item faz parte do segundo e quando o primeiro item é organizado (espacial, temporal e ativamente) com os outros componentes, de modo que juntos produzam o segundo item. Assim, o mecanismo da memória espacial possui vários níveis, alguns dos quais incluem órgãos tais como o hipocampo, que geram um mapa espacial, outros que envolvem interações celulares subjacentes à geração de mapas, e outros ainda que envolvem os mecanismos moleculares que subjazem aquelas interações celulares (CRAVER, 2007). Para mais informações sobre níveis, consulte a Seção 4.2.

2.4.6 Mecanismos Estáveis e Efêmeros

Finalmente, os mecanicistas acharam necessário distinguir entre mecanismos estáveis, que dependem fundamentalmente do arranjo mais ou menos fixo de partes e atividades, e mecanismos efêmeros, que envolvem um processo que evolui ao longo do tempo sem arranjo espacial e temporal fixos (GLENNAN, 2009). O mecanismo de controle de tempo em um relógio, por exemplo, é uma montagem relativamente estável de componentes em locais relativamente fixos que funcionam da mesma maneira, com os mesmos recursos organizacionais, cada vez que ele funciona. Os mecanismos efêmeros, por outro lado, envolvem um tipo de organização muito mais flexível: os itens ainda interagem no espaço e no tempo, mas não o fazem em virtude de estruturas estáveis e robustas. Muitos mecanismos químicos em uma célula são assim (RICHARDSON; STEPHAN, 2007). Mecanismos efêmeros são certamente um foco principal das ciências históricas, como arqueologia, história e biologia evolutiva (GLENNAN, 2009).

2.5 O Que Mecanismos Não São e O Que Não São Mecanismos

O termo "mecanismo" tem sido usado de muitas maneiras diferentes para expressar muitas ideias diferentes. A apropriação do termo pelos novos mecanicistas é, portanto, suscetível a associações inúteis, e a liberalização do termo por parte deles pode de levantar preocupações de que a noção de mecanismo tenha sido banalizada (*vide* MOSS, 2012; NICHOLSON, 2012). Agora, primeiro distinguiremos o novo mecanismo de outras doutrinas com as quais ele compartilha tanto nome quanto semelhança de família. Em seguida, discutiremos algumas coisas às quais o novo uso do termo "mecanismo" não se aplica.

2.5.1 O Que Mecanismos Não São

Os novos mecanicistas evitaram explicitamente as seguintes associações com o termo "mecanismo":

> 1 Mecanismos não são necessariamente **determinísticos**. Os mecanismos podem ser

estocásticos se, por exemplo, forem compostos de atividades estocásticas (BOGEN, 2005, 2008a) ou, em um sentido mais ordinário (isto é, consistente com o determinismo), sempre é possível que um ou mais fatores interfiram com o funcionamento de um mecanismo; uma das peças pode estar quebrada ou um obstáculo inesperado pode interferir na operação de um mecanismo. A verdade ou falsidade do determinismo, e sua relevância para a compreensão das ciências especiais, é uma questão independente da questão de saber se algo é um mecanismo.

2 O apelo a mecanismos não é necessariamente **reducionista**. Os mecanismos são frequentemente descritos como multiníveis, com atividades em diferentes níveis sendo igualmente essenciais para o funcionamento de um mecanismo. As explicações mecanísticas podem aparecer, acima, abaixo ou ao redor, dependendo da escolha de um *explanandum* e dos pressupostos do contexto explicativo (Bechtel 2009a). Mecanicistas podem ser reducionistas ou anti-reducionistas. Dito isto, muitos mecanicistas optam por alguma forma de anti-reducionismo explicativo, enfatizando a importância de explicações multiníveis e ascendentes, sem rejeitar as ideias centrais que motivam uma representação de mundo amplamente fisicalista (MCCAULEY; BECHTEL, 2001; CRAVER, 2007). Para uma discussão mais aprofundada, consulte Theurer (2013) e consulte, também, as Seções 3.1 e 5.

3 Nem todos os mecanismos são **máquinas**. Máquinas são artefatos criados pelo homem, com cada parte adicionada e organizada por alguém que o projeta para desempenhar uma função; os mecanismos biológicos e sociais, ao contrário, são produtos da evolução, em sentido amplo (DARDEN, 2006), e, portanto, exibem formas complexas de organização em comparação com os artefatos. Uma máquina pode conter vários mecanismos (um carro, por exemplo, possui mecanismos de frenagem,

propulsão, reprodução de música e controle climático). Máquinas também são capazes de serem ativas e passivas (um relógio parado ainda é uma máquina); mecanismos, por outro lado, têm um aspecto produtivo e estão sempre fazendo alguma coisa.

4 Os mecanismos não são necessariamente **sequenciais** ou **lineares**. Mecanismos podem ter ciclos de retroalimentação e ciclos em que o output do mecanismo ou componentes influencia, por sua vez, o *input* do mecanismo ou componentes em uma iteração subsequente (BECHTEL, 2011). Além disso, as interações entre componentes em um mecanismo não precisam ser descritas por uma equação linear.

5 Os mecanismos não são necessariamente **localizáveis** (BECHTEL; RICHARDSON, 2010). Os componentes dos mecanismos podem ser amplamente distribuídos (como muitos mecanismos cerebrais) e podem violar nosso senso intuitivo ou ensinado dos limites dos objetos (como um potencial de ação viola os limites das células). A presunção de localização é frequentemente uma heurística importante na busca por mecanismos; no entanto, essa heurística geralmente deve ser abandonada à medida que a organização do mecanismo se revela.

6 Os mecanismos não se limitam à **dinâmica *push-pull*** (fatores extrínsecos-fatores intrínsecos). O mecanismo de Descartes tinha essa característica, mas (como observado acima) o novo mecanismo explicitamente liberaliza a noção para justificar outros tipos de causas.

7 Mecanismos não são apenas **ficções/metáforas**. Quando uma cientista diz que existe um mecanismo que produz proteínas em organismos vivos, ela não está apenas usando a metáfora da máquina; antes, ela está dizendo que, de fato, existem partes e atividades organizadas nos organismos vivos de modo que elas produzam proteínas.

2.5.2 O Que Não São Mecanismos

Seria possível alegar que não resta nada do mecanismo, uma vez que ele perde essas associações históricas. É possível suspeitar que seu conceito tenha sido banalizado (DUPRÉ, 2013).

A ideia de mecanismo é uma parte central do ideal explicativo de entender o mundo ao aprender sua estrutura causal. A história da ciência contém muitas outras concepções de explicação e conhecimento científico que estão em desacordo com essa implicação. Uns sustentam que o mundo deve ser entendido em termos de motivos divinos. Outros sustentam que fenômenos naturais devem ser entendidos teleologicamente. E outros estão convencidos de que entender o mundo natural nada mais é do que ser capaz de prever seu comportamento. O compromisso com o mecanismo como um conceito estrutural é o compromisso com algo distinto, e para muitos, exclusivo, dessas concepções alternativas. Se isso parece trivial, e não uma conquista central na história da ciência, é porque a perspectiva mecanística agora domina completamente nossa visão de mundo científica.

No entanto, existem muitas maneiras de organizar os fenômenos além de revelar mecanismos. Alguns cientistas se preocupam com estruturas físicas e suas relações espaciais sem considerar como eles funcionam: um anatomista pode estar interessado na organização espacial de partes dentro do corpo com um interesse mínimo em como essas partes se articulam para produzir alguma coisa. Muitos cientistas constroem modelos preditivos de sistemas sem nenhuma pretensão de que esses modelos de fato revelem as estruturas causais pelas quais os sistemas funcionam.

Alguns cientistas estão preocupados com a taxonomia, classificando semelhantes com semelhantes sem considerar como os itens classificados surgiram ou como eles funcionam. Finalmente, em muitas áreas da ciência, existe uma distinção amplamente reconhecida e pragmaticamente significativa entre saber que **C** (por exemplo, fumar) é uma causa de **E** (câncer de pulmão) e saber como **C** causa **E**. Isso não é bem uma diferença ontológica, mas uma diferença no detalhe com o qual se procura entender a estrutura causal de um sistema. Em suma, existem muitos conceitos estruturais na ciência e nem todos podem ser assimilados a mecanismos.

Mas, o crítico poderia insistir, o que não conta como um mecanismo? Aqui estão algumas categorias de contraste:

9 **Entidades** (ou **objetos**) não são mecanismos. Mecanismos fazem coisas. Se um objeto não está

fazendo nada (isto é, se não há fenômeno), não é um mecanismo.

1 0**Correlações** não são mecanismos. Mecanismos explicam muitas correlações, e muitas correlações podem ser usadas para caracterizar relações causais ou mecanísticas, mas as correlações em si não são mecanísticas. O mesmo pode ser dito das meras sequências temporais de eventos.

1 1**Inferências**, **raciocínios** e **argumentos** não são mecanismos. Embora existam mecanismos de inferência e raciocínio, o que faz de algo uma inferência ou um raciocínio é uma relação lógica e não (meramente) uma relação causal entre premissa e conclusão.

1 2**Simetrias** não são mecanismos. Muitos tipos de simetria são de fundamental importância em diferentes áreas da física (por exemplo, simetrias translacionais, simetrias rotacionais). Essas são características dos sistemas físicos que são fatos ou suposições altamente gerais, e não mecanismos.

1 3**Leis fundamentais** e **relações causais fundamentais** não são mecanismos. Se uma lei ou relação causal é fundamental, então (por definição) não há mecanismo para ela.

1 4**Relações de necessidade lógica e matemática não são mecanismos**. Tais verdades são válidas em todos os mundos possíveis e assim não dependem, para a sua verdade, de fatos sobre a estrutura causal desse mundo.

Esta não é uma lista exaustiva de não-mecanismos ou de conceitos estruturais não-mecanísticos. No entanto, demonstra que mesmo o conceito liberalizado de mecanismo não é vazio nem trivial.

2.6 Trabalho Filosófico a Ser Feito

Grande parte da nova filosofia mecanicista se concentrou nas ciências especiais, como a neurociência e a biologia molecular. Nos anos seguintes, os filósofos estenderam a concepção mecanística para outras disciplinas científicas, como Biologia Celular (BECHTEL, 2006), Ciência Cognitiva (BECHTEL, 2008; THAGARD, 2006), Neuroeconomia (CRAVER; ALEXANDROVA, 2008), Química Orgânica (RAMSEY, 2008), Física (TELLER 2010), Astrofísica (ILLARI; WILLIAMSON, 2012), Genética do Comportamento (TABERY, 2014a) e filogenética (MATTHEWS, no prelo). Filósofos continuam testando os limites dessa abordagem, com a expectativa de que estruturas organizacionais alternativas possam desempenhar papéis centrais em outras ciências. Por exemplo, surgiu um debate na filosofia da biologia sobre se a seleção natural é ou não um mecanismo (*vide* SKIPPER; MILLSTEIN, 2005; BAKER, 2005; BARROS, 2008; BARROS, 2008; ILLARI; WILLIAMSON, 2012; HAVSTAD, 2011; MATTHEWSON; CALCOTT, 2011). Debates semelhantes têm surgido a respeito da explicação mecanística na ciência cognitiva (BECHTEL, 2008; PICCININI; CRAVER, 2011; WEISKOPF, 2011; POVICH, no prelo).

A iniciativa de compreensão dos mecanismos computacionais é uma área que tem recebido atenção especial. Em algumas abordagens, os mecanismos computacionais formam uma subclasse própria de mecanismos que podem ser definidos explicitamente em termos dos tipos de entidades, propriedades e atividades envolvidas nos mecanismos naquela classe (PICCININI, 2007; MILKOWSKI, 2013). De acordo com essa perspectiva, mecanismos computacionais são mecanismos que têm a função de manipular veículos de mídia-independente, de acordo com uma regra geral que se aplica a todos os veículos e depende dos inputs para sua aplicação (PICCININI; SCARANTINO, 2011). Os computadores digitais são distintos porque seus veículos são dígitos (PICCININI, 2007). Os proponentes dessa abordagem esperam diferenciar mecanismos computacionais de mecanismos não computacionais apelando para os componentes característicos pertencentes aos mecanismos da computação. Essa visão contrasta tanto com uma visão semântica, segundo a qual a computação é essencialmente uma questão de manipular símbolos ou representações, quanto com visões perspectivistas, de acordo com as quais um mecanismo conta como algo que computa se for descrito desta maneira (CHURCHLAND, 1986; CHURCHLAND; SEJNOWSKI, 1992; SHAGRIR, 2010).

Filósofos das ciências sociais também enfatizaram e debateram a importância do conhecimento mecanística (ELSTER, 1989). Para uma revisão pertinente dessas

conexões, consulte Hedström e Ylikoski (2010). Nesse contexto, os apelos aos mecanismos visam remediar o caráter relativo de desinformação das explicações sociais (ou a nível macro) dos fenômenos sociais (como normas generalizadas, desigualdades persistentes, redes e estruturas institucionais), insistindo que essas explicações sejam fundamentadas em detalhes mecanísticos sobre agentes e atores individuais, seus desejos e motivações e, principalmente, suas relações uns com os outros. A ênfase nas relações entre atores distancia essa visão mecanística do individualismo metodológico. Os mecanicistas das ciências sociais também tendem a se esquivar de teorias grandes e abrangentes, mas se dirigem a explicações mais locais: o conhecimento científico cresce adicionando itens a uma caixa de ferramentas composta por mecanismos, e mostrando como os itens dessa caixa de ferramentas podem ser combinados para fornecer uma explicação para um fenômeno particular. Frederica Russo (2009) discute várias estratégias para a modelagem de mecanismos sociais (LITTLE, 1991, 1998; HEDSTRÖM, 2005; HEDSTRÖM; SWEDBERG, 1998).

3. Explicação: Das Análises Formais às Estruturas Materiais

O modelo de explicação de lei de cobertura [*covering-law*] foi uma peça central da concepção de ciência lógico-empirista. De acordo com esse modelo, explicações são argumentos que mostram que o evento a ser explicado (o evento *explanandum*) deveria ser esperado com base nas leis da natureza e nas condições iniciais e limitantes (os *explanans*). Para os defensores do modelo de lei de cobertura, o problema filosófico da explicação é, em grande parte, uma questão de analisar a estrutura formal dos argumentos explicativos (HEMPEL; OPPENHEIM, 1948; HEMPEL, 1965). Um arco-íris, por exemplo, é explicado sob o modelo de lei de cobertura por referência a leis de reflexão e refração em conjunto com condições relativas à posição do sol e à natureza da luz, à posição das gotas de chuva e à posição da pessoa que vê o arco-íris. A descrição do arco-íris é a conclusão de um argumento dedutivo com afirmações nomológicas e descrições de condições como premissas; portanto, o arco-íris deveria ser esperado à luz do conhecimento das leis e condições.

Os mecanicistas, ao contrário, insistem que a explicação é uma questão de elucidar as estruturas causais que produzem, subjazem ou mantêm o fenômeno de interesse. Para os mecanicistas, o problema filosófico diz respeito principalmente à caracterização ou descrição das estruturas materiais ou ônticas às quais os

modelos explicativos (incluindo argumentos) devem referir para que sejam considerados genuinamente explicativos. Um arco-íris, para o mecanicista, é explicado situando esse fenômeno na estrutura causal do mundo; a explicação é uma descrição de como o fenômeno foi produzido por entidades (como gotas de chuva e globos oculares) com propriedades particulares (como formas e índices de refração) que interagem causalmente com a luz propagada pelo sol. Os mecanicistas normalmente distinguem várias maneiras de situar um fenômeno dentro da estrutura causal do mundo.

3.1 Explicações Etiológicas e Constitutivas

A maioria dos mecanicistas reconhece dois aspectos principais da explicação mecanística: o etiológico e o constitutivo. Salmon (1984) os descreve como duas maneiras diferentes de situar um fenômeno *explanandum* no nexo causal (CRAVER, 2001b; GLENNAN, 2009). Explicações etiológicas revelam a história causal do fenômeno *explanandum*, como quando se diz que um vírus explica uma doença. As explicações constitutivas, ao contrário, explicam um fenômeno descrevendo o mecanismo subjacente a ele, como quando se diz que regiões cerebrais, músculos e articulações explicam o ato de alcançar um objeto.

Os argumentos filosóficos contra o modelo de lei de cobertura focaram frequentemente em sua incapacidade de lidar com explicações causais e etiológicas. O modelo falhou em responder corretamente a uma variedade de casos problemáticos justamente porque tentou fornecer uma abordagem da explicação sem nenhuma menção explícita à causa (BROMBERGER, 1966; SALMON, 1984; SCRIVEN, 1959).

Os novos mecanicistas estendem esse tipo de crítica ao modelo de lei de cobertura da micro-redução interteórica. De acordo com o modelo de lei de cobertura da explicação reducionista, uma teoria sobre partes reduz, e assim explica, uma teoria sobre as totalidades quando é possível derivar a segunda da primeira supondo leis-pontes [*bridge laws*] para conectar as duas (NAGEL, 1961; Schaffner, 1993).

Alguns mecanicistas argumentam que o modelo de lei de cobertura de explicações constitutivas tem problemas análogos aos que afligem o modelo de lei de cobertura de explicações etiológicas. Os potenciais de ação não podem ser explicados por meras sequências temporais de eventos totalmente irrelevantes para o fenômeno, mas pode-se derivar uma descrição do potencial de ação a partir de descrições de tais fenômenos irrelevantes. Os potenciais de ação não podem ser explicados por meros padrões de correlação que não são indicativos de uma relação

causal subjacente. Subprodutos irrelevantes de um mecanismo podem estar correlacionados com o comportamento do mecanismo, tão perfeitamente correlacionados ao ponto de que se possam formar leis-ponte entre os níveis, mas sem que expliquem por meio disso o relacionamento. Simplesmente encontrar um correlato neural da consciência, por exemplo, não constituiria, e não é considerado por ninguém, como constituindo uma explicação da consciência. Assim, os mecanicistas argumentam que explicações micro-redutivas devem satisfazer restrições causais da mesma forma que explicações etiológicas (CRAVER, 2007).

O modelo de lei de cobertura também falha em distinguir modelos que meramente re-descrevem o fenômeno em termos gerais, de explicações que, além de prever aspectos do fenômeno, revelam os mecanismos que o produzem (CRAVER, 2006; KAPLAN; CRAVER, 2011; WEISKOPF 2011). Por exemplo, a lei de Snell permite prever como a luz se curva ao passar de um meio para outro, mas não explica por que a luz se curva. Os novos mecanicistas também argumentam que o modelo de lei de cobertura falha em distinguir modelos preditivamente adequados, mas ficcionais, de modelos explicativos. Finalmente, os mecanicistas argumentam que o modelo interteórico de redução falha em capturar uma dimensão importante da qualidade explicativa: a profundidade. Uma implicação do modelo de lei de cobertura é que quaisquer enunciados de lei verdadeiros que permitam derivar a lei *explanandum* (com correções e suposições adequadas) contarão como uma explicação completa. Contudo, parece que se pode aprofundar uma explicação abrindo caixas-pretas [*black boxes*] e revelando como as coisas funcionam até o nível que se tome como relativamente fundamental para os propósitos em questão. Tais críticas sugerem que o modelo de lei de cobertura da explicação constitutiva é muito fraco para capturar as normas de explicação nas ciências especiais.

Outros mecanicistas argumentaram que o modelo de lei de cobertura é muito forte. Os filósofos da biologia há muito defendem que não existem leis do tipo descrito pelo empiricista lógico na biologia e em outras ciências especiais (BEATTY, 1995; MITCHELL, 1997, 2000; WOODWARD, 2001). Pode-se concluir com isso que não há explicações na biologia (ROSENBERG, 1985), mas tal conclusão radical é dificilmente compatível com avanços óbvios em compreender, por exemplo, a síntese de proteínas, os potenciais de ação, a sinalização celular, e uma série de outros fenômenos biológicos. Em tais casos, verifica-se que os cientistas recorrem a mecanismos para realizar o trabalho explicativo, mesmo nos casos em que nada semelhante a uma lei pareça estar disponível.

3.2 Relevância Constitutiva

Ao redobrar a atenção à explicação constitutiva, os mecanicistas perceberam a necessidade de uma análise da relevância constitutiva, ponto fundamental para separar fatores relevantes de irrelevantes em um mecanismo (CRAVER, 2007; YLIKOSKI, 2013). Um sistema (**S**) exibindo o fenômeno (ψ) é composto por muitas entidades (**x**) diferentes, com várias propriedades, envolvidas em inúmeras atividades (φ) organizadas em conjunto (veja a Figura 1, acima, na Seção 2). Um problema de pesquisa central é dizer quais dessas entidades, atividades e características organizacionais contribuem para o fenômeno e quais não contribuem. Em certo sentido, esse é um desafio de definir as fronteiras de um mecanismo: de dizer o que está e o que não está no mecanismo.

Três propostas foram consideradas. A primeira, a **abordagem da manipulabilidade mútua**, entende a relevância constitutiva em termos das manipulações experimentais usadas para testar as relações interníveis. De acordo com esta abordagem, se é possível demonstrar (i) que os supostos componentes estão contidos em **S**, (ii) que algumas intervenções ideais no suposto componente (φ de **x**) mudam o fenômeno (ψ de **S**), e (iii) que algumas intervenções ideais no ψ de **S** mudam as φ de **x**, será suficiente para estabelecer que **x** é um componente do mecanismo. A noção de uma intervenção ideal nesta abordagem é explicitamente tributária, e uma proposta de extensão, da teoria da relevância causal de Woodward para o domínio constitutivo (CRAVER, 2007; KAPLAN, 2012). Uma preocupação com a abordagem da manipulabilidade mútua, entretanto, é que ela é mais um guia epistêmico para a relevância constitutiva, e não uma abordagem do que é relevância constitutiva (COUCH, 2011). A abordagem oferece, na melhor das hipóteses, uma condição suficiente de relevância. Além disso, a noção de uma intervenção "ideal", emprestada da abordagem de Woodward de relevância causal, não pode ser aplicada diretamente a explicações constitutivas. Uma intervenção **ideal** em um sistema não pode intervir na variável independente e na dependente ao mesmo tempo. No entanto, quando alguém intervém para produzir **S** ψ (ou impedir ψ de **S**), invariavelmente também intervém sobre os componentes do ψ de **S**. E quando alguém intervém nos componentes do ψ de **S**, frequentemente intervém no ψ de **S**. Como φ de **x** e ψ de **S** estão relacionados como parte do todo, eles não são independentes e, portanto, requerem outra maneira de pensar sobre intervenções ideais (*vide* BAUMGARTNER, 2010; 2013; LEURIDAN, 2011). Confira, ainda, Menzies (2012) e Woodward (2014).

Uma segunda proposta oferece uma abordagem de regularidade da relevância constitutiva inspirada na noção de Mackie de entender uma causa como uma condição INUS: uma parte Insuficiente, mas não redundante de uma condição não-necessária, mas suficiente para o efeito em questão (MACKIE, 1974; CUMMINS, 1983). Nessa abordagem, um componente constitutivamente relevante é uma parte insuficiente, mas não redundante de um mecanismo não-necessário, mas suficiente, para um determinado fenômeno (COUCH, 2011; HARBECKE, 2010, 2014). Permita que qualquer número de mecanismos possa ser suficiente para dar origem ao ψ de **S**; cada possível mecanismo suficiente é então não-necessário para ψ. Cada um desses mecanismos é feito de componentes, nenhum dos quais é por si só suficiente para produzir o comportamento do mecanismo como um todo, mas cada um deles é necessário no contexto do mecanismo para **S** to ψ. Esta abordagem pressupõe a ideia de "ser necessário no contexto", e pode-se razoavelmente se preocupar em separar **Xs** acidentalmente correlacionados, de **Xs** que de fato fazem uma diferença para ψ de **S**.

Uma terceira abordagem para a relevância constitutiva dispensa a concepção interníveis imposta pela abordagem da manipulabilidade mútua e tenta analisar a relevância usando apenas noções causais. De acordo com abordagens desse tipo, a relevância constitutiva é um tipo de intermediação [*between-ness*] causal. Se o ψ de **S** é entendido como uma relação de *input-output* de algum tipo, então a relevância mecanística pode ser entendida como sendo um elo necessário na cadeia causal entre o *input* e o *output* (HARINEN, no prelo; MENZIES, 2012). Os experimentos supostamente intraníveis na abordagem da manipulabilidade mútua podem então ser reformulados como diferentes tipos de experimentos causais de nível único [*unilevel*]. Romero (no prelo) fornece uma interessante perspectiva dessas questões e oferece a nova sugestão de que intervenções supostamente de alto nível são, na verdade, intervenções *fat-handed* (que manipulam diversas variáveis ao mesmo tempo) em relação às suas contrapartes de nível inferior.

3.3 Mecanismos e Modelos

A literatura filosófica sobre mecanismos também se intersecciona com a literatura filosófica sobre modelos científicos (*vide* verbete sobre **Modelos na Ciência**). Aqui distinguimos modelos mecânicos de modelos de mecanismos e discutimos variedades de modelos não mecânicos.

3.3.1 Caracterizando a Completude

Glennan (2005) propôs uma definição de **modelo mecânico** da seguinte forma:

> (MM) Um modelo mecânico consiste em (i) uma descrição do comportamento do mecanismo (a descrição comportamental); e (ii) uma descrição do mecanismo responsável por esse comportamento (a descrição mecânica) (p. 446).

Tais modelos podem ser representados de muitas maneiras diferentes (*vide* GIERE, 2004). Eles são avaliados em termos de sua capacidade de prever as características do fenômeno e em termos do mapeamento entre os itens no modelo e as entidades, atividades e características organizacionais no mecanismo (GLENNAN, 2005, p. 17; KAPLAN; CRAVER, 2011). Glennan enfatiza que não existe uma linha rígida que separa modelos completos e incompletos; em vez disso, os modelos estão continuamente em processo de articulação e refinamento. Considerações pragmáticas determinam se um modelo é completo o suficiente.

Este último ponto está relacionado à distinção de Darden entre **esquemas de mecanismo** e **esboços de mecanismo** (DARDEN; CAIN, 1989; DARDEN, 2002). Ao descobrir um mecanismo, muitas vezes é crucial identificar lacunas que devem ser preenchidas em um modelo. Mesmo que nenhum modelo esteja completo em um sentido absoluto, alguns modelos têm lacunas que devem ser preenchidas antes que estejam completos o suficiente.

Esquemas de mecanismo são descrições abstratas de mecanismos que podem ser preenchidas com detalhes para produzir um mecanismo tipo ou espécime específico. Assim, o esquema

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Proteína}$$

pode ser preenchido com uma sequência específica de bases no DNA, seu complemento no RNA e uma sequência de aminoácidos correspondente na proteína. As setas podem ser preenchidas, mostrando como a transcrição e a tradução funcionam. Um **esboço** de mecanismo é uma representação incompleta de um mecanismo que especifica algumas das entidades, atividades e características

organizacionais relevantes, mas deixa lacunas que ainda não podem ser preenchidas. Caixas pretas, pontos de interrogação e termos de preenchimento (como "ativar", "causar" ou "inibidor") reservam o lugar para alguma entidade, atividade ou processo ainda a ser descoberto. A distinção entre esboços e esquemas é uma questão de completude: esquemas são mais completos do que esboços no sentido de que um esboço omite um ou mais estágios do mecanismo que tem que ser entendidos a fim de resolver o problema da descoberta por completo.

Os novos mecanicistas também enfatizam a distinção entre um esquema (possível) **como-possivelmente** e um esquema ***como-verdadeiramente*** (verdadeiramente suficiente) (CRAVER; DARDEN, 2013). Um esquema como-possivelmente descreve como entidades e atividades **podem** ser organizadas para produzir um fenômeno. Um esquema como-possivelmente é uma hipótese n sobre como o mecanismo funciona. Tais esquemas podem ser verdadeiros (suficiente) ou falsos. Um esquema como-possivelmente verdadeiro (suficiente) é (embora possamos não saber) também um esquema como-verdadeiramente. Um esquema como-verdadeiramente descreve como entidades e atividades são de fato organizadas para produzir o fenômeno. O termo "como-verdadeiramente (suficiente)" captura a ideia de que o requisito "acurácia" de um esquema mecanístico pode variar consideravelmente de um contexto pragmático para outro (WEISBERG, 2013). Um esquema como-possivelmente falso é meramente um modelo como-possivelmente; *just-so-stories* são apenas modelos como-possivelmente (DRAY, 1957; BRANDON, 1985). Usado dessa forma, o termo "esquema como-possivelmente" é semelhante ao termo "hipótese": é cogitado como uma possibilidade, mas não necessariamente endossado.

3.3.2 Abstração e Idealização

Em contraste com esquemas e esboços de mecanismo, alguns modelos de mecanismos funcionam não descrevendo todas as partes, interações causais e características organizacionais, mas sim abstraindo tais detalhes potencialmente complicadores (CRAVER; DARDEN, 2013; STREVENS, 2008; LEVY; BECHTEL, 2012). Em tais casos, suposições idealizadoras podem ser introduzidas para destacar a característica relevante do mecanismo: populações infinitas, planos sem atrito, formas geométricas perfeitas, são presumidos a fim de despojar o modelo de detalhes que não importam, ou que apenas obstruiriam, as finalidades pretendidas do modelo.

Críticos da nova filosofia mecânica insistiram na importância da abstração na ciência, chamando a atenção para as discussões acima sobre completude. As metas de completude e acurácia são consideradas conflitantes com a prática comum de utilização de modelos que sacrificam detalhes e verdade em troca de clareza e generalidade (STREVENS, 2008; WOODWARD, 2014). A distinção normativa entre um esquema e um esboço, por exemplo, parece sugerir que a ciência progride passando de modelos incompletos a modelos completos. E a distinção entre como-possivelmente e como-realmente-suficiente parece, da mesma, forma privilegiar a precisão sobre outras metas de modelagem, que muitas vezes requerem distorção e falsidade (*vide* WIMSATT 2007; WEISBERG, 2007; LEVY; BECHTEL, 2012; BATTERMAN; RICE, 2014; CHIRIMUUTA, 2014; LEVY, 2014).

No entanto, os mecanicistas podem certamente permitir que nem todos os modelos de mecanismos sejam modelos mecânicos ou esquemas de mecanismo. Frequentemente outros tipos de modelo são úteis para isolar aspectos centrais do funcionamento de um mecanismo. Modelos dinâmicos, por exemplo, podem ser usados para caracterizar a dinâmica temporal de um mecanismo (BECHTEL, 2013a, 2013b; KAPLAN; BECHTEL, 2011). Modelos de rede podem ser usados para caracterizar padrões de conectividade independentemente de quais unidades estão conectadas e independentemente de quais tipos de conexões se está particularmente interessado em caracterizar (HUNNEMAN, 2010). Modelos mínimos podem ser usados para capturar algo fundamental sobre a dinâmica de uma ampla classe de mecanismos que não compartilham entidades e atividades em comum (BATTERMAN, 2002). Um modelo de um mecanismo é um modelo que descreve um mecanismo. Não precisa ser um modelo mecânico ou um esquema de mecanismo, no sentido acima, para desempenhar esse papel.

Alguns mecanicistas reservam o termo "modelos mecânicos" para modelos que descrevem as entidades, atividades e características organizacionais de um sistema. De acordo com a abordagem de Glennan (2005), um modelo mecânico que deixa de fora algumas características relevantes é*, ipso facto*, incompleto e esquemático. Uma exemplificação específica desse debate diz respeito à força explicativa dos modelos funcionais em psicologia. Piccinini e Craver (2011) argumentam que tais modelos devem ser entendidos como esboços mecanísticos, modelos caixa-preta a serem avaliados e preenchidos conforme detalhes sobre o mecanismo subjacente são descobertos. Modelos caixa-preta são incompletos em virtude de omitir detalhes sobre os mecanismos subjacentes e de que esses modelos, em última instância, dependem, para sua força explicativa, da promessa de que os

modelos funcionais correspondem, de fato, ao modo como o mecanismo funciona. Confira Weiskopf (2011) para uma crítica a este relato e, ainda, Povich (no prelo) para uma resposta.

Pode-se falar sobre a explicação mecanística abstraindo o tipo de modelo usado para descrever o mecanismo: o compromisso com a explicação mecanística não é um compromisso sobre a forma do modelo, mas sim um compromisso sobre o que tais modelos devem representar: a saber, estruturas causais e mecanísticas. Modelos são explicativos em virtude do fato de que representam as estruturas causais/mecanísticas que produzem, subjazem ou mantêm o fenômeno. Eles são não-mecanísticas se se referem a algum tipo de relação não-causal e não-mecanística (SALMON, 1984; CRAVER, 2014).

3.4 Trabalho Filosófico a Ser Feito

Até o momento, muito do trabalho sobre a explicação mecanística tem sido impulsionado pelo objetivo de fornecer uma teoria da explicação mecanística descritiva e normativamente adequada. Alguns afirmam que existem tipos de explicação que dependem muito pouco de uma compreensão precisa dos detalhes mecanísticos de um sistema (WOODWARD, 2014; WEISKOPF, 2011) ou que funcionam fundamentalmente removendo todos esses detalhes de um modelo (BATTERMAN; RICE, 2014). Resolver esses debates exigirá muita clareza sobre exatamente o que se espera de uma teoria filosófica da explicação científica e o que se considera uma explicação científica (STREVENS, 2008; CRAVER, 2014). Investigações serão necessárias para se compreender as diversas formas representacionais que os cientistas usam para representar mecanismos (BURNSTON *et al*., no prelo) e para compreender o papel da idealização na explicação mecanística (LEVY; BECHTEL, 2012; HUNEMAN, 2010). Trabalhos adicionais também são necessários para delinear as fronteiras entre a explicação mecanística e outras supostas variedades de explicação, e para dizer, de forma tão perspicaz quanto Hempel ou quanto a teoria causal-mecânica, o que um modelo deve fazer para contar como explicativo e precisamente como as boas explicações devem ser distinguidas das ruins.

4. Metafísica dos Mecanismos

Nesta seção, revisamos algumas das maneiras como o conceito de mecanismo tem sido usado em diversas áreas da metafísica. De todas as áreas que discutimos, esta, provavelmente, é a que mais necessita de desenvolvimentos futuros. Aqui, discutimos a relação entre mecanismos e leis, emergência, realização, tipos naturais e funções.

4.1 Mecanismos de Instância/Espécime, Mecanismos de Tipo e Leis

Em grande parte da literatura inicial sobre mecanismos, eles são contrastados explicitamente com as leis da natureza (BECHTEL, 1988; BECHTEL; ABRAHAMSEN, 2005; MDC, 2000). Esse contraste surgiu claramente de um consenso emergente na filosofia de que existem poucas, ou talvez nenhuma, leis da biologia (*vide* Seção 3.1). As generalizações empíricas que encontramos na biologia tendem a ser limitadas por cláusulas *ceteris paribus*; se elas são ou não válidas depende das condições de base que podem não ser válidas e das condições internas do mecanismo que podem falhar. Essas generalizações, em resumo, são explicáveis mecanicamente; a necessidade delas deriva de um mecanismo (CUMMINS, 2000; GLENNAN, 1996). Mecanismos, portanto, parecem desempenhar o papel de leis nas ciências biológicas: buscamos mecanismos para explicar, prever e controlar fenômenos na natureza, mesmo se os mecanismos carecem de muitas das características definitivas das leis na abordagem empirista lógica (como universalidade, necessidade inviolável, ou escopo irrestrito).

Uma vertente específica desta discussão emergiu da consideração do argumento de Weber (2005) de que a biologia é **heterônima**, ou seja, que em última análise toma emprestado seu poder explicativo das leis da física e da química. Weber usa o modelo de Hodgkin e Huxley do potencial de ação como um exemplo da redução dos fenômenos biológicos às leis físicas (como a lei de Ohm e a equação de Nernst). Craver (2006) responde que a força explicativa do modelo de Hodgkin e Huxley, de fato, requer uma compreensão das propriedades biológicas distintas dos canais iônicos, cujas propriedades foram ocultadas [*black-boxed*] na equação de corrente total de Hodgkin e Huxley (*vide* CRAVER, 2007; BOGEN, 2008b; WEBER, 2008).

No entanto, o contraste entre leis e mecanismos nem sempre foi totalmente claro. Alguns, como Bogen (2005), Machamer (2004) e Glennan (no prelo) enfatizam

que as causas e mecanismos são, no fundo, singulares, não gerais ou universais. Leuridan (2010), baseando-se no trabalho de Mitchell (2000), alega que os mecanismos não podem substituir as leis da natureza em nossa compreensão conceitual da explicação e da metafísica da ciência. Os cientistas raramente investigam mecanismos de instância/espécime, pode-se pensar, mas estão muito mais interessados em tipos. E quando começamos a falar sobre tipos de mecanismos, voltamos ao assunto de formular regularidades gerais sobre como eles funcionam. Assim, parece que o conceito de mecanismo não substitui o trabalho que a generalização deveria fazer, mas requer a ideia de regularidade e, portanto, algo semelhante a leis, se for para fazer o trabalho explicativo (vide ANDERSEN 2011, 2012, 2014a, 2014b; KRICKEL, 2014). Para uma resposta a Leuridan, consulte Kaiser e Craver (2013).

4.2 Mecanismos, Níveis e Emergência

O desenvolvimento de mecanismos também ajudou a esclarecer a ideia sobre níveis de organização e sua relação com outras formas de organização e formas não-mecanísticas de emergência.

Muitos mecanicistas, seguindo Simon (1996), enfatizam que os sistemas biológicos são hierarquicamente organizados em estruturas quase decomponíveis: mecanismos dentro de mecanismos, dentro do mecanismo. Usando a parábola de *Tempus* e Hora, Simon (1962) argumentou que um relojoeiro que fabrica relógios hierarquicamente decomponíveis (*Tempus*) fará mais relógios do que aquele que fabrica relógios holísticos (Hora). Esta parábola levou Simon à conclusão de que as estruturas evoluídas são mais prováveis de serem quase decomponíveis em estruturas e subestruturas hierarquicamente organizadas, mais ou menos estáveis. Alguns autores alegaram que a história é enganosa, porque a evolução não constrói organismos do zero, peça por peça (BECHTEL, 2009b). Steel (2008), baseando-se no trabalho de outros (SCHLOSSER; WAGNER, 2004) tenta reconstruir este argumento, portanto, como uma forma de mostrar que sistemas resultantes de evolução são mais prováveis de serem modulares: sistemas feitos de partes independentemente manipuláveis podem colocar em quarentena os efeitos de alterações em partes específicas, o que lhes dá uma flexibilidade adicional para fazer alterações locais sem causar efeitos colaterais catastróficos.

A decomposição aproximada dos mecanismos está diretamente relacionada à ideia de que os mecanismos abrangem vários níveis de organização. O comportamento

do todo é explicado em termos das atividades e interações entre as partes componentes. Essas atividades e interações são elas mesmas sustentadas por atividades e interações subjacentes entre partes componentes e assim por diante (*vide* BECHTEL; RICHARDSON, 2010). Craver (2007) define níveis de mecanismos em termos de uma relação entre o comportamento (ψ) exibido por um sistema (**S**) e a atividade (φ) de alguma parte componente (**X**) desse sistema. Nessa abordagem, a atividade de uma parte do sistema (φ de **X**) está em um nível inferior de organização mecanística do que o comportamento do sistema (ψ de **S**) se e somente se (i) **X** é uma parte de **S**, e (ii) φ de **X** é um componente em ψ de **S**. Em suma, dizer que algo está em um nível mecanístico inferior ao mecanismo como um todo é dizer que ele é uma parte operante do mecanismo. Embora o termo "nível" seja usado de muitas maneiras legítimas, níveis de mecanismos parecem desempenhar um papel central na estruturação das relações entre muitos modelos diferentes na biologia contemporânea, por exemplo, entre a genética mendeliana e molecular (DARDEN, 2006), entre a aprendizagem e a memória e fisiologia do canal (CRAVER, 2007), e entre a variação no nível de população e mecanismos de desenvolvimento (TABERY, 2009, 2014a).

Uma implicação dessa visão em níveis, combinada com certas suposições familiares sobre relações causais, é que não pode haver relações causais entre itens em diferentes níveis de mecanismos. Pode haver relações causais entre coisas de diferentes tamanhos, e pode haver relações causais entre coisas descritas em vocabulários muito diferentes; mas (novamente, associadas a certas suposições sobre a assimetria temporal de causa e efeito e a independência de causa e efeito) não pode haver relações causais entre o comportamento de um mecanismo e as atividades das partes que conjuntamente constituem aquele comportamento. Afirmações sobre causalidade intranível, que são onipresentes na literatura científica, são mais bem compreendidas como almejando um sentido diferente de níveis ou, no que diz respeito aos níveis de mecanismos, como expressando afirmações híbridas que combinam afirmações constitutivas sobre a relação entre o comportamento do mecanismo como um todo e as atividades de suas partes, e afirmações causais acerca de relações entre coisas não relacionadas como parte e todo (CRAVER; BECHTEL, 2007). Para interpretações alternativas sobre níveis, consulte Fehr (2004), Leuridan (2011), Thalos (2013), Eronen (2013, 2015), Baumgartner e Gebhardter (no prelo) e Romero (no prelo). Para reflexões sobre o status metafísico de fenômenos de nível superior e causas de nível superior, consulte Baumgartner (2010), Glennan (2010a, 2010b) e Hoffman-Kolss (2014).

Conforme observado acima, o fato de que os fenômenos em níveis superiores de mecanismos dependem da organização das partes componentes implica que as propriedades/atividades do todo não são simples somas das propriedades/atividades das partes. Níveis de mecanismos podem, portanto, ser contrastados com níveis de mera agregação. Como o todo é maior (neste sentido) do que a soma das partes, alguns (como Wimsatt) consideraram apropriado descrevê-lo como um tipo de emergência. A **emergência mecanística** (*ou* **organizacional**) assim entendida é onipresente e banal, mas extremamente importante para compreender como os cientistas explicam as coisas.

Igualmente familiar é a **emergência epistêmica**, a incapacidade de prever as propriedades ou comportamentos do todo a partir de propriedades e comportamentos das partes. A emergência epistêmica pode surgir como resultado da ignorância, como não reconhecer uma variável relevante, ou por não saber como diferentes variáveis interagem em redes complexas. Isso também pode resultar de limitações nas habilidades cognitivas humanas ou nas ferramentas representacionais da geração atual (BEDAU, 1997; BOOGERD, 2005; RICHARDSON; STEPHAN, 2007). A necessidade prática de estudar mecanismos decompondo-os em partes componentes levanta o desafio epistêmico de juntar as partes novamente de uma maneira que realmente funcione (BECHTEL, 2013a).

A ênfase dos mecanicistas na emergência mecanística/organizacional e epistêmica contrasta com seu desejo de se distanciarem da emergência fantasmagórica [*spooky*] (RICHARDSON; STEPHAN, 2007). A emergência fantasmagórica envolveria o aparecimento de novas propriedades nos mecanismos sem uma base suficiente. Não está claro se as propriedades emergentes são adequadamente entendidas como propriedades dos mecanismos necessários; e não está claro em que sentido a propriedade emergente é "emergente", ao invés de uma característica fundamental da estrutura causal do mundo.

Em suma, tais formas de emergência são completamente distintas, e, portanto, não ganham plausibilidade se associadas verbalmente com a emergência organizacional/mecanística e epistêmica.

4.3 Mecanismos e Realização

Como o conceito estrutural de um mecanismo é tão útil para pensar sobre níveis e explicação nas ciências, alguns estudiosos buscaram na noção de mecanismo

uma maneira de dar corpo à relação ontológica de realização.

De acordo com a "concepção plana" (GILLETT, 2002), realização é uma relação entre diferentes propriedades de uma mesma coisa (KIM, 1998; SHAPIRO, 2000; SHOEMAKER, 2003, 2007; POLGER, 2007). A concepção de subconjunto, que sustenta que uma propriedade P1 (por exemplo, energia cinética média do gás) produz a propriedade P2 (por exemplo, temperatura do gás) quando os poderes causais distintivos de P2 (temperatura) são um subconjunto dos poderes causais distintivos de P2 (energia cinética média), é um exemplo da visão plana. P1 e P2 são atribuídos à mesma coisa, o gás (GILLETT, 2002, 2003). A concepção dimensionada descreve a realização como uma relação mantida entre as propriedades do todo e as propriedades das partes e sua organização. Essa concepção de realização condiz com os objetivos explicativos das ciências especiais e se encaixa perfeitamente com a base evidencial na qual as afirmações intraníveis são fundamentadas (*vide* AIZAWA; GILLETT, 2011). Gillett (2013) expandiu esta noção para lidar com a realização de objetos, propriedades e processos. Para críticas e alternativas, consulte Polger (2010), Melnyk (2003) e Melnyk (2010).

4.4 Mecanismos e Tipos Naturais

Teorias mecanísticas de tipos naturais se desenvolveram a partir da concepção de **aglomerado homeostático de propriedades** [*homeostatic property cluster* ou HPC] de Boyd. A concepção HPC é uma teoria de tipos naturais projetada para funcionar em domínios com alta variabilidade individual. A concepção HPC é oferecida como uma terceira via entre o essencialismo e o nominalismo sobre tipos nas ciências especiais (BOYD, 1991, 1997, 1999; KORNBLITH, 1993; WILSON, 1999, 2005).

De acordo com essa concepção, um tipo natural é caracterizado por i) um aglomerado de propriedades que co-ocorrem regularmente, e ii) um mecanismo de geração de similaridade que explica por que as propriedades em (i) tendem a co-ocorrer. Em resumo, tipos são aglomerados de propriedades explicados por mecanismos.

Esta visão dos tipos naturais tem sido empregada para argumentar a favor de uma revisão taxonômica, por exemplo, da biologia da emoção humana (GRIFFITHS, 1997), da estrutura de conceitos (MACHERY, 2009) e da taxonomia da psiquiatria (KENDLER; ZACHAR; CRAVER, 2010; *vide* CRAVER, 2009): um tipo supostamente único é dividido em múltiplos tipos porque propriedades distintas em um aglomerado

de propriedades são explicadas por mecanismos distintos. Essa visão de tipos também pode ser usada para compreender os tipos que são historicamente transitórios e, de alguma maneira, produto de atitudes humanas e, portanto, socialmente construídos neste sentido direto; talvez o conceito de raça pode ser entendido dessa maneira (*vide* KUORIKOSKI; PÖYHÖNIN, 2012; KHALIDI, 2013).

4.5 Mecanismos e Funções

A ênfase na importância dos mecanismos está historicamente associada à rejeição da teleologia e das causas formais (WESTFALL, 1971). No entanto, a biologia contemporânea e muitas outras ciências especiais, apesar da aceitação generalizada do conceito estrutural de mecanismo, continuam a fazer uso do conceito de função, uma noção teleológica (consulte o verbete sobre **noções teleológicas em biologia**). Como a noção de função em jogo na ciência contemporânea está relacionada ao conceito de mecanismo? Craver (2001a), seguindo Cummins (1975), argumenta que a descrição funcional é um meio em perspectiva de situar alguma parte dentro de um mecanismo de nível superior. De acordo com essa visão, a teleologia não é uma característica do mundo, mas é imposta a ele por um descritor intencional (*vide* MACHAMER, 1977). Garson (2011, 2012), seguindo Wimsatt (1972b), Wright (1973) e Neander (1991a, b), argumenta que funções são efeitos de um item, que fazem parte da explicação etiológica (por meio de seleção, aprendizagem ou reforço) sobre o porquê de o item estar presente; como tal, as funções são reduzidas a histórias causais. Em uma terceira visão, Maley e Piccinini (no prelo) argumentam que a função (teleológica) de um item é sua contribuição para as finalidades dos organismos, que podem ser finalidades objetivas ou subjetivas. Da mesma maneira que Garson, Maley e Piccinini afirmam que as funções são objetivas (isto é, não relativas ao observador). Ao contrário de Garson, no entanto, eles não se baseiam na etiologia do item, mas em sua contribuição atual para a sobrevivência ou reprodução (as finalidades objetivas dos organismos) ou para o que o próprio organismo deseja (as finalidades subjetivas dos organismos).

4.6 Trabalho Filosófico a Ser Feito

Como deve ser o mundo para que essa perspectiva mecanística seja precisa? Claramente, há muitas maneiras de responder a essa pergunta a partir de diferentes suposições metafísicas iniciais. E claramente, muitas dessas suposições descartam esta imagem do mundo como ilegítima. O caminho mais claro a seguir, ao que parece, é descobrir precisamente os comprometimentos associados ao sustentar que o mundo é composto de uma hierarquia de mecanismos e o que pode ser entendido com base em diferentes suposições iniciais.

Dito isso, nem todas as aplicações da estrutura mecanicista exigem uma metafísica completamente articulada. O trabalho de descoberta e explicação pode prosseguir perfeitamente bem sem abraçar qualquer imagem metafísica do mundo em particular. Filósofos com distintos interesses (descoberta, explicação, teste, redução, emergência, e assim por diante) tendem a elaborar o conceito de maneiras diferentes. Há várias razões para duvidar que a ideia de mecanismo possa receber uma única análise metafísica que abordará adequadamente os diversos fins filosóficos para os quais o conceito está sendo empregado.

5. Relações entre Disciplinas Científicas: Da Redução da Teoria à Integração do Mecanismo

5.1 Redução da Teoria

De acordo com Nagel (1961), a redução é uma espécie de explicação da lei de cobertura [*covering-law*]: uma teoria é reduzida a outra quando é possível identificar os termos teóricos da primeira com os da segunda, e literalmente derivar a primeira da segunda. Partindo do pressuposto de que disciplinas e teorias científicas correspondem umas às outras, a redução serve também como um modelo de integração interdisciplinar. As objeções do mecanicista ao modelo de lei de cobertura da explicação constitutiva (ou seja, micro-redução) são discutidas acima (*vide* Seção 3.1); aqui, o foco está em como distintas disciplinas da ciência são integradas.

Na visão de Nagel, a redução é um relacionamento intraníveis. É também uma relação entre teorias. Teorias sobre fenômenos em um nível superior (por exemplo, gases, relâmpagos e vida) são reduzidas a (isto é, derivadas de) teorias

sobre fenômenos em níveis inferiores (por exemplo, moléculas, elétrons e sistemas fisiológicos). Finalmente, a relação é formalmente especificada e tem pouco a ver com o conteúdo das teorias ou com as estruturas materiais que essas teorias descrevem. Da perspectiva mecanística, cada uma dessas características do modelo de Nagel é problemática.

5.2 Integração do Mecanismo

Primeiro, os mecanicistas criticam a ideia de que a redução deve ser entendida primordialmente como uma relação entre **teorias**. Ao integrar seus resultados, os cientistas não estão simplesmente construindo teorias em sentido absoluto; eles estão construindo teorias sobre mecanismos. Mecanismos podem ser talvez descritos usando abordagens formais de teorias – talvez eles possam ser axiomatizados usando lógica de predicados ou reconstruídos como predicados da teoria de conjuntos. Mas essas abordagens formais das estruturas das teorias científicas escondem as estruturas mecanísticas cruciais para a compreensão de como essas teorias são construídas e avaliadas (CRAVER, 2001b).

Os mecanicistas também questionam a ideia de que disciplinas se relacionam por meio da relação entre suas teorias. Darden e Maull (1977) argumentaram que os campos disciplinares muitas vezes integram suas descobertas por meio da construção de teorias intercampos, apelando para diversas relações materiais entre itens nos domínios dos diferentes campos: relações como causa e efeito, parte e todo, ou estrutura e função. Darden e Maull não ofereceram uma abordagem geral das teorias intercampos, mas Bechtel sugeriu profeticamente que tais teorias frequentemente assumem a forma de descrições de mecanismos (BECHTEL, 1988, p. 101-102).

A abordagem mecanística também tem sido apontada como tendo muitas vantagens comparada a redução em como pensar sobre as formas de integração interdisciplinar intraníveis. Primeiramente, ela fornece uma maneira direta de interpretar o debate de níveis (*vide* Seções 2.4.5 e 4.2). Em segundo lugar, ela oferece consideravelmente mais insights sobre o que é integração intraníveis, sobre as restrições (em termos de evidência) pelas quais as pontes intraníveis são avaliadas, e sobre as forças que impulsionam a coevolução do trabalho em diferentes níveis. Restrições nas partes, suas interações causais, e sua organização espacial, temporal e hierárquica ajudam a definir uma integração intraníveis. Finalmente, os mecanicistas reconhecem repetidamente a necessidade de não apenas se olhar

para baixo para os mecanismos constitutivos responsáveis por um determinado fenômeno (enfatizado por modelos de redução clássicos), mas também olhar para cima e ao redor para o contexto no qual o fenômeno está incorporado: a integração entre níveis é um esforço para enxergar como fenômenos em muitos níveis diferentes estão relacionados entre si (BECHTEL, 2009a; CRAVER, 2007).

Os mecanicistas desenvolveram vários exemplos estendidos das muitas formas de integração de mecanismos buscadas em programas de pesquisa mecanísticas. Darden (2005), por exemplo, sugere que os filósofos dominados pela redução clássica entenderam fundamentalmente mal a relação entre a genética mendeliana e a molecular. Embora os reducionistas vejam isso como um exemplo de explicação intranível, ela argumenta, isso é de fato um caso em que diferentes cientistas trabalharam em diferentes partes de um mecanismo que são etiologicamente (não constitutivamente) relacionados entre si. A genética mendeliana não se reduziu à biologia molecular; ao contrário, geneticistas clássicos e biólogos moleculares integraram seu trabalho concentrando-se em diferentes entidades de funcionamento nos mecanismos cromossômicos que operam sequencialmente e nos mecanismos moleculares hereditários. Exemplos também foram extraídos da descoberta dos mecanismos de síntese proteica (DARDEN, 2006) e biologia celular (BECHTEL, 2006). Craver (2007) usa exemplos da neurociência da memória para explorar como a integração multinível procede e deve proceder. Em cada caso, a busca por mecanismos serve como uma plataforma abstrata para a qual, e em torno da qual, convergem as descobertas de diversos cientistas.

5.2.1 Pluralismo Integrativo

A perspectiva mecanística tende a enfatizar o pluralismo integrativo na pesquisa científica (MITCHELL, 2003, 2009). O objetivo não é explicar o menos fundamental em termos do mais fundamental em uma relação gradual entre as teorias monolíticas em um nível e as teorias monolíticas em outro. Em vez disso, essas conquistas científicas são colaborativas e fragmentadas, adicionando restrições progressivas a uma imagem emergente de como um mecanismo funciona tanto em um nível quanto entre níveis. As muitas disciplinas científicas que investigam um fenômeno coexistem pluralisticamente e se co-informam mutuamente, contribuindo integrativamente para as explicações mecanísticas etiológicas, constitutivas e contextuais desse fenômeno (BECHTEL, 2009a; TABERY, 2014a).

5.3 Trabalho filosófico a ser feito

O modelo de Nagel de redução da teoria oferece uma visão clara da "unidade da ciência". De acordo com o modelo, a unidade entre as disciplinas científicas é alcançada reduzindo as teorias das disciplinas de nível superior às teorias das disciplinas de nível inferior. A integração, nessa visão, é entendida como um progresso em direção a um grandioso corpo unificado de conhecimento científico. Para os mecanicistas, em contraste, a integração é gradual, local e pluralista. Que tipo de unidade essa "integração" poderia sustentar? Essa questão se desenrola em uma discussão entre Longino e Tabery a respeito das relações disciplinares nas ciências do comportamento. Tabery argumenta que disciplinas tão díspares quanto neurobiologia e genética quantitativa podem coexistir pluralisticamente e co-informar as explicações causais de comportamentos complexos umas das outras por meio de integração de mecanismo. Longino rebate que o pluralismo de Tabery é apenas um tipo "moderado", porque o impulso para a integração, em última análise, é um impulso para a unificação (LONGINO, 2013, 2014; TABERY, 2014a, b). Sullivan (2009) também questiona a pressão para a integração do mecanismo; ela argumenta que existem barreiras significativas para o tipo de integração que os mecanicistas vislumbram. Diferentes laboratórios usam protocolos experimentais diferentes para estudar o que presumem ser o mesmo fenômeno; no entanto, esses diferentes protocolos, de fato, focam frequentemente em fenômenos diferentes, de modo que a integração alcançada pela combinação de resultados é apenas ilusória. Essas discussões são sintomáticas de questões filosóficas mais gerais enfrentadas pelos mecanicistas: Como as integrações de mecanismo são realmente alcançadas (em oposição a apenas afirmadas)? E qual é a relação entre integração e unificação do mecanismo? A nova filosofia mecânica se beneficiará de esforços futuros para situar a integração mecanística em visões filosóficas mais gerais de integração e pluralismo.

6. Descoberta: Dos Momentos Eureka às Estratégias de Descoberta

6.1 Descoberta via Momentos Eureka

O que os filósofos podem dizer sobre a descoberta científica? Muitos empiricistas lógicos tinham uma resposta simples: Nada. De acordo com Popper,

por exemplo, os filósofos podem elucidar a epistemologia do teste, mas não podem dizer nada de substancial sobre como os cientistas geram as ideias a serem testadas (POPPER, 1959). Tais momentos "eureka!", de criatividade, pertencem à esfera da Psicologia, não da Filosofia. Reichenbach distinguiu o contexto de descoberta do contexto de justificação (a "distinção de contexto") (REICHENBACH, 1938). Confira o verbete sobre **Hans Reichenbach** para uma interpretação alternativa dessa distinção. O processo de descoberta científica estava, portanto, em grande parte interditado aos filósofos.

Nem todos os filósofos da ciência concordaram. Hanson, por exemplo, articulou uma lógica de descoberta envolvendo inferências abdutivas de dados anômalos para novas hipóteses projetadas para explicá-los (HANSON, 1958). Outros focaram em metodologias de descoberta que poderiam permitir reconstruir racionalmente por que algo era descoberto em um determinado momento (NICKELS, 1985) ou explicar por que uma nova hipótese é considerada promissora e digna de posterior investigação (SCHAFFNER, 1993). As primeiras contribuições para a nova filosofia mecânica seguiram esse caminho e caracterizaram as estratégias investigativas que os cientistas usam para descobrir mecanismos (*vide* verbete sobre **Descoberta Científica**).

6.2 Descoberta via Estratégias

A obra de Bechtel e Richardson, *Discovering Complexity*, de 1993 (2010), está organizada em torno de um fluxograma que representa pontos de escolha na descoberta de um mecanismo. O processo de busca de mecanismos começa com uma caracterização provisória do fenômeno. Depois, segue as estratégias de **localizar** o mecanismo dentro do sistema e **decompor** o fenômeno em subfunções distintas. A localização da função envolve determinar quais dessas subfunções do sistema são desempenhadas por quais partes. Bechtel e Richardson caracterizam ainda o uso de experimentos excitatórios e inibitórios para obter esses tipos de informação. Bechtel e Abrahamsen (2013) acrescentam uma etapa subsequente, na qual os cientistas **recompõem** o que aprenderam sobre as partes funcionais reconstruindo-as para produzir o fenômeno em questão (talvez usando simulações).

Darden também enfatizou os mecanismos como um importante conceito estrutural na descoberta científica (DARDEN, 1980, 1982, 1986, 1991). Na descoberta da síntese proteica (investigada conjuntamente por biólogos moleculares e bioquímicos nas décadas de 1950 e 1960), os cientistas não tiveram simplesmente um momento

"eureka". Em vez disso, eles empregaram estratégias para revelar como um mecanismo funciona (DARDEN, 2006; CRAVER; DARDEN, 2013). Darden caracteriza o processo de descoberta do mecanismo como um "processo estendido e gradativo com hipóteses submetidas a refinamento iterativo"; esse processo ocorre por meio da construção, avaliação e revisão de esquemas de mecanismo à luz de restrições observacionais e experimentais (DARDEN, 2006, p. 272).

As estratégias de construção de Darden são estratégias para gerar novas hipóteses sobre um mecanismo. Além da decomposição e localização, Darden mostra que os cientistas muitas vezes **tomam emprestado um tipo de esquema** de outra área da ciência, como quando os mecanismos do tipo de seleção foram emprestados para entender como o sistema imunológico funciona, ou montam um mecanismo a partir de módulos conhecidos de atividade funcional (**submontagem modular**), como é comum em bioquímica e biologia molecular. Às vezes, os cientistas conhecem uma parte do mecanismo e tentam trabalhar para frente ou para trás nas outras partes e atividades. Na descoberta do mecanismo de síntese proteica, por exemplo, os biólogos moleculares trabalharam de trás para frente, a partir da estrutura do DNA, para descobrir quais moléculas poderiam interagir com ele (**encadeamento progressivo**), e os bioquímicos trabalharam de frente para trás, a partir das proteínas, para descobrir quais seriam as reações químicas necessárias para criá-las (**encadeamento reverso**). Eles se encontraram no meio no RNA. A síntese proteica agora envolve a transcrição de DNA em RNA e, em seguida, a tradução do RNA em proteínas. Longe de ser filosoficamente inescrutável, Darden aponta que os cientistas usaram o que sabiam sobre as entidades e atividades em funcionamento no mecanismo para inferir o que poderia vir ou a seguir ou antes no mecanismo de síntese proteica (DARDEN, 2006).

Estratégias de avaliação, para Darden, envolvem **raciocínio baseado em restrições** para delimitar os contornos do espaço de possíveis mecanismos para um dado fenômeno. Frequentemente, os cientistas raciocinam sobre como um mecanismo funciona se baseando em descobertas fundamentais sobre a organização espacial e temporal de suas partes. Harvey, por exemplo, raciocinou sobre a circulação do sangue considerando a localização das válvulas das veias e sua orientação em relação ao coração. Essas restrições organizacionais, e muitas outras, combinaram-se para estreitar o espaço dos mecanismos possíveis a uma pequena região contendo um modelo no qual o sangue completa um circuito do corpo (CRAVER; DARDEN, 2013).

Darden e Craver também discutem estratégias experimentais para aprender como um mecanismo funciona. Essas estratégias revelam como diferentes entidades e atividades em um mecanismo agem, interagem, e se organizam juntas. Por exemplo, alguém pode remover um suposto componente para ver se e como o mecanismo funciona na sua ausência (experimentos **inibitórios**). Ou pode-se **estimular** esse componente para ver se ele pode acionar o mecanismo ou modular seu comportamento. Ou pode-se **ativar** um mecanismo colocando-o nas condições precipitantes do fenômeno e observar como a entidade ou atividade muda conforme o mecanismo funciona. Craver (2002) discute isso sob o título de "experimentos intraníveis" (*vide* HARINEN, no prelo). Craver e Darden (2013) também discutem tipos mais complexos de experimentos para conhecer que tipo de entidade ou atividade contribui para um processo e para conhecer características mais complexas da organização de um mecanismo.

Datteri (2009; DATTERI; TAMBURRINI, 2007), explora o uso de simulações robóticas para fins de teste de mecanismos. Eles discutem como as suposições são construídas em modelos robóticos e como os experimentos podem ser projetados para revelar como os mecanismos funcionam. Este trabalho amplia a estrutura mecanística para a área da bio-robótica e revela um conjunto de estratégias distintas daquelas exploradas no trabalho de Darden.

Em vez de focar no processo pelo qual os esquemas de mecanismo são construídos, avaliados e revisados, Steele foca na questão de como se extrapola a partir de uma amostra de população ou um organismo modelo para a estrutura de um mecanismo em questão. Um tratamento comprovado para suprimir tumores em camundongos (um organismo modelo) também vai servir para suprimir tumores em humanos (a população-alvo)? Depois de desenvolver uma abordagem probabilística dos mecanismos, Steele considera como os pesquisadores se esquivam do que ele chama de círculo do extrapolador: determinar

> como poderíamos saber que o modelo e o alvo são semelhantes em aspectos causalmente relevantes antes de conhecer a relação causal no alvo (STEEL, 2008, p. 78)

Steel quebra o círculo do extrapolador ao desenvolver uma estratégia de extrapolação baseada em mecanismos – a estratégia de **rastreamento de processo comparativo**. Uma vez que um mecanismo para algum fenômeno tenha sido

elucidado em um modelo (como um processo particular de carcinogênese em ratos), os cientistas (toxicologistas neste caso), então, comparam as etapas cruciais (particularmente as etapas consecutivas) do modelo com as etapas do alvo, prestando atenção especial aos pontos do processo nos quais as diferenças são mais prováveis de surgir. Quanto maiores as semelhanças das entidades, atividades e organização dos mecanismos em ambas as populações, mais forte é a base para extrapolação; quanto maiores as diferenças, mais fraca é a base (vide HOWICK *et al*., 2013).

6.3 Evidência Mecanística nas Descobertas Médicas

A descoberta na medicina é outro âmbito onde a filosofia mecânica foi aplicada. Thagard baseia-se no caso da *H. pylori* como causa de úlceras para dar uma explicação de como os mecanismos de investigação contribuem para a descoberta científica.

Thagard chama a atenção tanto para a evidência estatística que sugere que as úlceras estão de alguma maneira associadas ao *H. pylori*, quanto para a evidência mecanística que pode explicar como o agente da infecção poderia persistir em um ambiente hostil por tempo suficiente para causar uma úlcera. Mais recentemente, filósofos interessados na medicina baseada em evidências têm sondado a relação entre esses dois tipos de evidências nas ciências da saúde. Russo e Williamson argumentam que ambos os tipos de evidência são necessários para justificar a inferência causal; a evidência correlacional estabelece que há uma relação que faz diferença [*difference-making*] entre alguma causa e algum efeito, enquanto a evidência mecanística estabelece como exatamente a causa produz seu efeito – a "Tese de Russo-Williamson" (RUSSO; WILLIAMSON, 2007). Filósofos, desde então, refinaram a Tese de Russo-Williamson, apontando, por exemplo, que "tipo de evidência" poderia se referir a diferentes metodologias de coleta de evidência ou a diferentes objetos de evidência. As metodologias de *difference-making* incluem estudos observacionais e ensaios clínicos randomizados, enquanto as metodologias mecanísticas incluem experimentos intervencionistas tais como os descritos acima; da mesma forma, o objeto da evidência poderia ser a evidência de uma diferença associada ou poderia ser a evidência relativa ao mecanismo que conecta causa e efeito (ILLARI, 2011; *vide* CAMPANER, 2011). Hierarquias médicas baseada em evidências, que classificam diferentes tipos de evidência em termos de sua força epistêmica, tendem a priorizar evidências de metodologias *difference-making* (como ensaios clínicos randomizados e meta-análises) em detrimento de evidências mecanísticas; em resposta, esses filósofos argumentam que os diferentes tipos de evidência estão

em igualdade de condições (cada um com seus próprios pontos fortes e fracos), e defendem a integração entre *difference-making* e a evidência mecanística, um sentimento que se alinha com a ênfase na integração do mecanismo discutida na Seção 5.2 (CLARKE *et al*., 2013, 2014).

6.4 Trabalho Filosófico a Ser Feito

Muitos mecanicistas exploraram as estratégias que os cientistas usam na descoberta. Bechtel e Richardson dedicaram-se à decomposição e localização; Darden e Craver destacaram o encadeamento progressivo e reverso; Russo e Williamson enfatizaram tanto o recurso *difference-making* quanto a evidência mecanística. Essas estratégias foram encontradas em ciências experimentais específicas, tais como neurociência e biologia molecular. Portanto, uma tarefa para os filósofos que estão avançando neste tópico é avaliar se existem ou não estratégias semelhantes em outras ciências, especialmente naquelas que operam fora do laboratório tradicional, tanto nas ciências humanas (tais como sociologia e economia) e nas ciências físicas (tais como cosmologia).

Também esperamos um espantoso desenvolvimento como resultado do preenchimento da lacuna entre as abordagens qualitativas dos mecanismos e a explicação mecanística desenvolvida no novo mecanismo e as teorias quantitativas de descoberta oriundas da disciplina de aprendizado de máquina e modelagem causal (SPIRTES *et al*., 2000; PEARL, 2009). O último oferece ferramentas para extrair dados correlacionais para dependências causais. Tais ferramentas podem escapar de abordagens mais qualitativas e históricas e podem, de fato, ir além das estratégias comuns que os cientistas tradicionalmente usam. Tais ferramentas também oferecem um meio de avaliar as estratégias de descoberta ao explorar as condições sob as quais elas têm sucesso ou falham e a eficiência com que emitem vereditos acerca de hipóteses causais.

7. Conclusão

A nova filosofia mecanicista e, de maneira mais geral, a atenção ao conceito estrutural de "mecanismo" tem se expandido rapidamente nas últimas duas décadas trazendo consigo novas orientações com respeito a uma ampla gama de questões

na filosofia da ciência. No entanto, está claro que muitos dos principais tópicos estão apenas começando a se desenvolver, deixando muito trabalho para os estudiosos elaborarem os compromissos básicos dessa concepção e considerarem o que significa fazer ciência fora dessa concepção. É provável que no futuro próximo haja uma discussão contínua das implicações e limites do uso dessa concepção para se pensar sobre a ciência e a prática científica.

Referências

AIZAWA, K.; GILLETT, C. The Autonomy of Psychology in the Age of Neuroscience in Causality in the Sciences. *In:* ILLARI, P. M.; RUSSO, F.; WILLIAMSON, J. (ed.). **Causality in the Sciences**. Oxford: Oxford University Press, 2011, p. 202-223.

ALLEN, G.E. Mechanism, Vitalism and Organicism in Late Nineteenth and Twentieth Century Biology: The Importance of Historical Context. *In:* CRAVER, C. F.; DARDEN, L. **In Search of Mechanisms**: Discoveries Across the Life Sciences. Chicago: University of Chicago Press, 2013, p. 261–283.

ALON, U. **An Introduction to Systems Biology.** Boca Raton: Chapman and Hall/ CRC Press, 2006.

ANDERSEN, H. Mechanisms, Laws, and Regularities. **Philosophy of Science**, v. 78, n. 2, p. 325-331, 2011.

ANDERSEN, H. The Case for Regularity in Mechanistic Causal Explanation. **Synthese**, v. 189, n. 415-432, 2012.

ANDERSEN, H. A Field Guide to Mechanisms: Part I. **Philosophy Compass**, v. 4, p. 274-283, 2014a.

ANDERSEN, H. A Field Guide to Mechanisms: Part II. **Philosophy Compass**, v. 4, p. 283-297, 2014b.

BAKER, J. M. Adaptive Speciation: The Role of Natural Selection in Mechanisms of Geographic and Non-geographic Speciation. In: CRAVER, C. F.; DARDEN, L. **In Search of Mechanisms**: Discoveries Across the Life Sciences. Chicago: University of Chicago Press, 2013, p. 303-326.

BARROS, D. B. Natural Selection as a Mechanism. **Philosophy of Science**, v. 75, p. 306-322, 2008.

BATTERMAN, R. Asymptotics and the Role of Minimal Models. **The British Journal for the Philosophy of Science**, v. 53, p. 21-38, 2002.

BATTERMAN, R.; RICE, C. Minimal Model Explanations. **Philosophy of Science**, v. 81, p. 349-376, 2014.

BAUMGARTNER, M. Interventionism and Epiphenomenalism. **Canadian Journal of Philosophy**, v. 40, p. 359-384, 2010.

BAUMGARTNER, M. Rendering Interventionism and Non-Reductive Physicalism Compatible. **Dialectica**, v. 67, p. 1-27, 2013.

BAUMGARTNER, M.; GEBHARTER, A. Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. **The British Journal for the Philosophy of Science**. DOI:10.1093/bjps/axv003.

BEATTY, J. The Evolutionary Contingency Thesis. *In:* LENNOX, J. G.; WOLTERS, G. (ed.). **Concepts, Theories, and Rationality in the Biological Sciences.** Pittsburgh: University of Pittsburgh Press, 1995, p. 45-81.

BECHTEL, W. **Philosophy of science**: An overview for cognitive science. Hillsdale: Erlbaum, 1988.

BECHTEL, W. **Discovering Cell Mechanisms**: The Creation of Modern Cell Biology. Cambridge: Cambridge University Press, 2006.

BECHTEL, W. **Mental Mechanisms**: Philosophical Perspectives on Cognitive Neuroscience. London: Routledge, 2008.

BECHTEL, W. Looking Down, Around, and Up: Mechanistic Explanation in Psychology. **Philosophical Psychology**, v. 22, p. 543-564, 2009a.

BECHTEL, W. Explanation: Mechanism, Modularity, and Situated Cognition. *In:* ROBBINS, P.; AYDEDE, M. (ed.). **Cambridge handbook of situated cognition**. Cambridge: Cambridge University Press, 2009b, p. 155-170.

BECHTEL, W. Mechanism and Biological Explanation. **Philosophy of Science**, v. 78, p. 533-557, 2011.

BECHTEL, W. Addressing the Vitalist's Challenge to Mechanistic Science: Dynamic Mechanistic Explanation. *In:* NORMANDIN, S.; WOLFE, C. T. (ed.). **Vitalism and the Scientific Image in Post-Enlightenment Life Science, 1800–2010.** Dordrecht: Springer, 2013a, p. 345-370.

BECHTEL, W. From Molecules to Behavior and the Clinic: Integration in Chronobiology. **Studies in History and Philosophy of Biological and Biomedical Sciences**, v. 44, p. 493-502, 2013b.

BECHTEL, W.; ABRAHAMSEN, A. Explanation: A Mechanistic Alternative. **Studies in History and Philosophy of the Biological and Biomedical Sciences,** v. 36, p. 421-441, 2005.

BECHTEL, W.; ABRAHAMSEN, A. Thinking Dynamically about Biological Mechanisms: Networks of Coupled Oscillators. **Foundations of Science**, v. 18, p. 707-723, 2013.

BECHTEL, W.; RICHARDSON, R. C. **Discovering Complexity**: Decomposition and Localization as Strategies in Scientific Research. 2. ed. Cambridge: MIT Press/Bradford Books, 2010.

BEDAU, M. Weak Emergence. **Philosophical Perspectives**, v. 11, p. 375-399, 1997.

BEEBEE, H.; HITCHCOCK, C.; MENZIES, P. (ed.). **The Oxford Handbook of Causation**. Oxford: Oxford University Press, 2010.

BOAS, M. The Establishment of the Mechanical Philosophy. **Osiris**, v. 10, p. 412-541, 1952.

BOGEN, J. Regularities and Causality; Generalizations and Causal Explanations. *In:* CRAVER, C. F.; DARDEN, L. **In Search of Mechanisms**: Discoveries Across the Life Sciences. Chicago: University of Chicago Press, 2013, p. 397-420.

BOGEN, J. Causally Productive Activities. **Studies in History and Philosophy of Science**, v. 39, p. 112-123, 2008a.

BOGEN, J. The Hodgkin-Huxley Equations and the Concrete Model: Comments on Craver, Schaffner, and Weber. **Philosophy of Science**, v. 75, p. 1034-1046, 2008b.

BOOGERD, F. C.; BRUGGEMAN, F. J.; RICHARDSON, R. C.; STEPHAN, A.; WESTERHOFF, H. V. Emergence and Its Place in Nature: A Case Study of Biochemical Networks. **Synthese**, v. 145, p. 131-164, 2005.

BOYD, R. Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. **Philosophical Studies**, v. 61, p. 127-148, 1991.

BOYD, R. Kinds as the 'Workmanship of Men: Realism, Constructivism, and Natural Kinds. *In:* NIDA-RUMELIN, J. (ed.). **Rationality, Realism, Revision**: Proceedings of the 3rd International Congress of the Society for Analytical Philosophy. New York: Walter de Gruyter, 1997, p. 52-89.

BOYD, R. Homeostasis, Species, and Higher Taxa. *In:* WILSON, R. A. (ed.). **Species.** Cambridge: MIT Press,1999, p. 141-185.

BRANDON, R. Grene on Mechanism and Reductionism: More Than Just a Side Issue. *In:* ASQUITH, P.; KITCHER, P. (ed.). **PSA 1984.** v. 2. East Lansing: Philosophy of Science Association, 1985, p. 345-353.

BROMBERGER, S. Why Questions. *In:* COLODNY, R. G. (ed.). **Mind and Cosmos.** Pittsburgh: University of Pittsburgh Press, 1966, p. 86-111.

BURNSTON, D.C.; SHEREDOS, B.; ABRAHAMSEN, A.; BECHTEL, W. ´Scientists ´ Use of Diagrams in Developing Mechanistic Explanations: A Case Study from Chronobiology. **Pragmatics and Cognition**, v. 22, n. 2, p. 224-243, 2014.

CAMPANER, R. Understanding Mechanisms in the Health Sciences. **Theoretical Medicine and Bioethics**, v. 32, p. 5-17, 2011.

CARTWRIGHT, N. D. **Nature's Capacities and their Measurement**. New York: Oxford University Press, 1989.

CARTWRIGHT, N. D. **The Dappled World**: A Study of the Boundaries of Science. Cambridge: Cambridge University Press, 1999.

CARTWRIGHT, N. D. Modularity: It Can—and Generally Does—Fail. *In:* COSTANTINI, D.; GALAVOTTI, M. C.; SUPPES, P. (ed.). **Stochastic Dependence and Causality**. Stanford: CSLI Publications, 2001, p. 65-84.

CARTWRIGHT, N. D. Against Modularity, the Causal Markov Condition and Any Link Between the Two: Comments on Hausman and Woodward. **British Journal for the Philosophy of Science**, v. 53, p. 411-453, 2002.

CHEMERO A.; SILBERSTEIN, M. After the Philosophy of Mind: Replacing Scholasticism with Science. **Philosophy of Science**, v. 75, p. 1-27, 2008.

CHURCHLAND, P. S. **Neurophilosophy:** Toward a Unified Science of the Mind/ Brain. Cambridge: MIT Press, 1986,

CHURCHLAND, P.S.; SEJNOWSKI, T. J. **The Computational Brain**. Cambridge: MIT Press, 1992.

CHIRIMUUTA, M. Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience. **Synthese**, v. 191, p. 127-153, 2014.

CLARKE, B.; GILLIES, D.; ILLARI, P.; RUSSO, F.; WILLIAMSON, J. The Evidence that Evidence-Based Medicine Omits. **Preventive Medicine**, v. 57, p. 745-747, 2013.

CLARKE, B.; GILLIES, D.; ILLARI, P.; RUSSO, F.; WILLIAMSON, J. Mechanisms and the Evidence Hierarchy. **Topoi**, v. 33, p. 339-360, 2014.

COUCH, M. B. Mechanisms and Constitutive Relevance. **Synthese**, v. 183, p. 375-388, 2011.

CRAVER, C. F. Role Functions, Mechanisms and Hierarchy. **Philosophy of Science**, v. 68, p. 31-55, 2001a.

CRAVER, C. F. Structures of Scientific Theories. *In:* MACHAMER, P. K.; SILBERSTEIN, M. (ed.). **Blackwell Guide to the Philosophy of Science.** Blackwell: Oxford, 2001b, p. 55-79.

CRAVER, C. F. When Mechanistic Models Explain. **Synthese**, v. 153, p. 355-376, 2006.

CRAVER, C. F. **Explaining the Brain**: Mechanisms and the Mosaic Unity of Neuroscience. Oxford: Clarendon Press, 2007.

CRAVER, C. F. Mechanisms and Natural Kinds. **Philosophical Psychology**, v. 22, p. 575-594, 2009.

CRAVER, C. F. Functions and Mechanisms: A Perspectivalist Account. *In:* HUNEMAN, P. (ed.). **Functions:** Selection and Mechanisms. Dordrecht: Springer, 2013, p. 133-158.

CRAVER, C. F. The Ontic Conception of Scientific Explanation. *In:* HÜTTEMAN, A.; KAISER, M. (ed.). **Explanation in the Special Sciences**: The Cases of Biology and History. Dordrecht: Springer, 2014, p. 27-52.

CRAVER, C. F.; ALEXANDROVA, A. No Revolution Necessary: Neural Mechanisms for Economics. **Economics and Philosophy**, v. 24, p. 381-406, 2008.

CRAVER, C. F.; BECHTEL, W. M. Top-down Causation without Top-down Causes. **Biology and Philosophy**, v. 22, p. 547-563, 2007.

CRAVER, C. F.; DARDEN, L. **In Search of Mechanisms**: Discoveries Across the Life Sciences. Chicago: University of Chicago Press, 2013.

CUMMINS, R. Functional Analysis. **Journal of Philosophy**, v. 72, p. 741-764, 1975.

CUMMINS, R. **The Nature of Psychological Explanation.** Cambridge: Bradford/ MIT Press, 1983.

CUMMINS, R. How Does It Work? Vs. 'What Are The Laws? Two Conceptions of Psychological Explanation. *In:* KEIL, F.; WILSON, R. (ed.). **Explanation and Cognition.** Cambridge, MA: MIT Press, 2000, p. 117-144.

DARDEN, L. Reasoning in Theory Construction: Analogies, Interfield Connections, and Levels of Organization. *In:* WEINGARTNER, P.; DORN, G. (ed.). **Foundations of Biology.** Vienna: Holder-Pichler-Tempsky, 1986, p. 99-107.

DARDEN, L. **Theory Change in Science**: Strategies from Mendelian Genetics. New York: Oxford University Press, 1991.

DARDEN, L. Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining. **Philosophy of Science**, v. 69, p. S354-S365, 2002.

DARDEN, L. Relations Among Fields: Mendelian, Cytological and Molecular Mechanisms. **Studies in History and Philosophy of Biological and Biomedical Sciences**, v. 36, p. 349-371, 2005.

DARDEN, L. **Reasoning in Biological Discoveries**: Mechanism, Interfield Relations, and Anomaly Resolution. New York: Cambridge University Press, 2006.

DARDEN, L.; CAIN, J. Selection Type Theories. **Philosophy of Science**, v. 56, p. 106-129, 1989.

DARDEN, L.; MAULL, N. Interfield Theories. **Philosophy of Science**, v. 44, p. 43-64, 1977.

DATTERI, E. Simulation Experiments in Bionics: A Regulative Methodological Perspective. **Biology and Philosophy**, v. 24, p. 301-324, 2009.

DATTERI, E.; TAMBURRINI, G. Biorobotic Experiments for the Discovery of Biological Mechanisms. **Philosophy of Science**, v. 74, p. 409-430, 2007.

DESAUTELS, L. Against Regular and Irregular Characterizations of Mechanisms. **Philosophy of Science**, v. 78, p. 914-925, 2011.

DES CHENE, D. **Spirits & Clocks**: Machine & Organism in Descartes. Ithaca: Cornell University Press, 2001.

DES CHENE, D. Mechanisms of Life in the Seventeenth Century: Borelli, Perrault, Régis. **Studies in the History and Philosophy of Biological and Biomedical Sciences**, v. 36, p. 245-260, 2005.

DIJKSTERHUIS, E. J. **The Mechanization of the World Picture.** New York: Oxford University Press, 1961.

DOWE, P. Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. **Philosophy of Science**, v. 59, p. 195-216, 1992.

DOWE, P. The Causal-Process-Model Theory of Mechanisms. *In:* ILLARI, P. M.; RUSSO, F.; WILLIAMSON, J. (ed.). **Causality in the Sciences**. Oxford: Oxford University Press, 2011, p. 865-879.

DRAY, W. **Laws and Explanation in History.** London: Oxford University Press, 1957.

DUPRÉ, J. **The Disorder of Things:** Metaphysical Foundations of the Disunity of Science. Cambridge: Harvard University Press, 1993.

DUPRÉ, J. Living Causes. **Proceedings of the Aristotelian Society**, v. 87, p. 19-35, 2013.

ELSTER, J. **Nuts and Bolts for the Social Sciences**. Cambridge: Cambridge University Press, 1989.

ERONEN, M. I. No Levels, No Problems: Downward Causation in Neuroscience. **Philosophy of Science**, v. 80, p. 1042-1052, 2013.

ERONEN, M. I. Levels of Organization: A Deflationary Account. **Biology and Philosophy**, v. 30, p. 39-58, 2015.

FAGAN, M. B. The Joint Account of Mechanistic Explanation. **Philosophy of Science**, v. 79, p. 448-472, 2012.

FAGAN, M. B. **Philosophy of Stem Cell Biology**: Knowledge in Flesh and Blood. London: Palgrave MacMillan, 2013.

FEHR, C. Feminism and Science: Mechanism without Weductionism. **National Women's Studies Association Journal**, v. 16, p. 136-156, 2004.

FODOR, J. **Psychological Explanation**. New York: Random House, 1968.

GARSON, J. Selected Effects Functions and Causal Role Functions in the Brain: The Case for an Etiological Approach to Neuroscience. **Biology & Philosophy**, v. 26, p. 547-565, 2011.

GARSON, J. Function, Selection, and Construction in the Brain. **Synthese**, v. 189, p. 451-481, 2012.

GARSON, J. The Functional Sense of Mechanism. **Philosophy of Science**, v. 80, p. 317-333, 2013.

GIERE, R.N. How Models Are Used to Represent Reality. **Philosophy of Science**, v. 71, p. 742-752, 2004.

GILLETT, C. The Dimensions of Realization: A Critique of the Standard View. **Analysis**, v. 62, p. 316-323, 2002.

GILLETT, C. The Metaphysics of Realization, Multiple Realizability, and the Special Sciences. **The Journal of Philosophy**, v. 100, p. 591-603, 2003.

GILLETT, C. Constitution, and Multiple Constitution, in the Sciences: Using the Neuron to Construct a Starting Framework. **Minds and Machines**, v. 23, p. 309-337, 2013.

GLENNAN, S. S. Mechanisms and The Nature of Causation. **Erkenntnis**, v. 44, p. 49-71, 1996.

GLENNAN, S. S. Capacities, Universality and Singularity. **Philosophy of Science**, v. 64, p. 605-626, 1997.

GLENNAN, S. S. Rethinking Mechanistic Explanation. **Philosophy of Science**, v. 69, p. S342-S353, 2002.

GLENNAN, S. S. Modeling Mechanisms. **Studies in the History and Philosophy of the Biological and Biomedical Sciences**, v. 36, p. 375-388, 2005.

GLENNAN, S. S. Productivity, Relevance and Natural Selection. **Biology and Philosophy**, v. 24, p. 325-339, 2009.

GLENNAN, S. S. Mechanisms. 2010 a. *In:* BEEBEE, H.; HITCHCOCK, C.; MENZIES, P. (ed.). **The Oxford Handbook of Causation**. Oxford: Oxford University Press, 2010, p. 315-325.

GLENNAN, S. S. Mechanisms, Causes, and the Layered Model of the World. **Philosophy and Phenomenological Research**, v. 81, p. 362-381, 2010b.

GLENNAN, S. S. **The New Mechanical Philosophy**. Oxford: Oxford University Press, 2017.

GODFREY-SMITH, P. Causal Pluralism. *In:* BEEBEE, H.; HITCHCOCK, C.; MENZIES, P. (ed.). **The Oxford Handbook of Causation**. Oxford: Oxford University Press, 2010, p. 326-337.

GOULD, S. J. **Wonderful Life**: Burgess Shale and the Nature of History. New York: W.W. Norton and Company, 1990.

GRIFFITHS, P. E. **What Emotions Really Are**: The Problem of Psychological Categories. Chicago: University of Chicago Press, 1997.

GRUSH, R. In Defense of Some 'Cartesian' Assumptions Concerning the Brain and its Operation. **Biology and Philosophy**, v. 18, p. 53-93, 2003.

HANSON, N. R. **Patterns of Discovery**. Cambridge: Cambridge University Press, 1958.

HARBECKE, J. Mechanistic Constitution in Neurobiological Explanations. **International Studies in the Philosophy of Science**, v. 24, p. 267-285, 2010.

HARBECKE, J. Regularity Constitution and the Location of Mechanistic Levels. **Foundations of Science**, v. 19, 2014. DOI: 10.1007/s10699–014–9371–1.

HARINEN, T. Mutual Manipulability and Causal Betweenness. **Synthese**, v. 195, p. 35-54, 2018. DOI:10.1007/s11229-014-0564-5.

HAUGELAND, J. **Having Thought**. Cambridge: Harvard University Press, 1998.

HAVSTAD, J. C. Discussion: Problems for Natural Selection as a Mechanism. **Philosophy of Science**, v. 78, p. 512-523, 2011.

HEDSTRÖM, P. **Dissecting the Social**: On the Principles of Analytical Sociology. Cambridge: Cambridge University Press, 2005.

HEDSTRÖM, P.; SWEDBERG, R. **Social Mechanisms**: An Analytical Approach to Social Theory. Cambridge: Cambridge University Press, 1998.

HEDSTRÖM, P.; YLIKOSKI, P. Causal Mechanisms in the Social Sciences. **Annual Review of Sociology**, v. 36, p. 49-67, 2010.

HEIL, J.; MELE, A. **Mental Causation**. Oxford: Clarendon Press, 1993.

HEMPEL, C. G. **Aspects of Scientific Explanation and Other Essays in the Philosophy of Science**. New York: Free Press, 1965.

HEMPEL, C. G.; OPPENHEIM, P. Studies in the Logic of Explanation. **Philosophy of Science**, v. 15, p. 135-175, 1948.

HITCHCOCK, C. R. Discussion: Salmon on Explanatory Relevance. **Philosophy of Science**, v. 62, p. 304-320, 1995.

HOFFMANN-KLOSS, V. Interventionism and Higher-Level Causation. **International Studies in the Philosophy of Science**, v. 28, p. 49-64, 2014.

HOWICK, J. P. G.; ARONSON, J. K. Problems with Using Mechanisms to Solve the Problem of Extrapolation. **Theoretical Medicine and Bioethics**, v. 34, p. 275-291, 2013.

HUNEMAN, P. Topological Explanations and Robustness in Biological Sciences. **Synthese**, v. 177, p. 213-245, 2010.

ILLARI, P. M. Disambiguating the Russo-Williamson Thesis. **International Studies in the Philosophy of Science**, v. 25, p. 139-157, 2011.

ILLARI, P. M.; RUSSO, F.; WILLIAMSON, J. (ed.). **Causality in the Sciences**. Oxford: Oxford University Press, 2011.

ILLARI, P. M.; WILLIAMSON, J. Function and Organization: Comparing the Mechanisms of Protein Synthesis and Natural Selection. **Studies in the History and Philosophy of the Biological and Biomedical Sciences**, v. 41, p. 279-291, 2010.

ILLARI, P. M.; WILLIAMSON, J. What is a Mechanism?: Thinking about Mechanisms Across the Sciences. **European Journal for Philosophy of Science**, v. 2, p. 119-135, 2012.

KAISER, M.; CRAVER, C. F. Mechanisms and Laws: Clarifying the Debate. *In:* CHAO, H-K.; CHEN, S-T.; MILLSTEIN, R. L. (ed.). **Mechanism and Causality in Biology and Economics**. Dordrecth: Springer, 2013, p. 125-145.

KAPLAN, D.M. How to Demarcate the Boundaries of Cognition. **Biology and Philosophy**, v. 27, p. 545-570, 2012.

KAPLAN, D. M.; BECHTEL, W. Dynamical Models: An Alternative or Complement to Mechanistic Explanations, Topics. **Cognitive Science**, v. 3, p. 438-444, 2011.

KAPLAN, D. M.; CRAVER, C. F. The Explanatory Force of Dynamical Models. **Philosophy of Science**, v. 78, p. 601-627, 2011.

KAUFFMAN, S. A. Articulation of Parts Explanation in Biology and the Rational Search for Them. *In:* BUCK, R. C.; COHEN, R. S. (ed.). **PSA 1970, Boston Studies in the Philosophy of Science.** v. 8. Dordrecht: Reidel, 1971, p. 257-272.

KENDLER, K., P. ZACHAR, CRAVER, C. F. What Kinds of Things are Psychiatric Disorders? **Psychological Medicine**, v. 41, p. 1143-1150, 2010.

KHALIDI, M. A. **Natural Categories and Human Kinds**. Cambridge: Cambridge University Press, 2013.

KIM, J. **Mind in a Physical World**. Cambridge: The MIT Press, 1998.

KRICKEL, B. **The Metaphysics of Mechanism, PhD Dissertation**. Berlin: Humboldt-Universität zu Berlin, 2014.

KORNBLITH, H. **Inductive Inference and Its Natural Ground**. Cambridge: MIT Press, 1993.

KUHLMANN, M. Mechanisms in Dynamically Complex Systems. *In:* ILLARI, P. M. Disambiguating the Russo-Williamson Thesis. **International Studies in the Philosophy of Science**, v. 25, 2011, p. 880-906.

KUHLMANN, M.; GLENNAN, S. On the Relation between Quantum Mechanical and Neo-Mechanistic Ontologies and Explanatory Strategies. **European Journal for Philosophy of Science**, v. 4, p. 337-359, 2014.

KUHN, T. S. **The Structure of Scientific Revolutions**. Chicago: University of Chicago Press, 1962.

KUORIKOSKI, J.; PÖYHÖNEN, S. Looping Kinds and Social Mechanisms. **Sociological Theory**, v. 30, p. 187-205, 2012.

KUORIKOSKI, J.; YLIKOSKI, P. K. **EPSA11 Perspectives and Foundational Problems in Philosophy of Science**. Heidelberg: Springer, p. 69-80, 2013.

LAKATOS, I. **The Methodology of Scientific Research Programmes**: Philosophical Papers. v. 1. Cambridge: Cambridge University Press, 1977.

LAUDAN, L. **Progress and Its Problems**. Berkeley: University of California Press, 1977.

LEURIDAN, B. Can Mechanisms Really Replace Laws of Nature? **Philosophy of Science**, v. 77, p. 317-340, 2010.

LEURIDAN, B. Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms. **The British Journal for the Philosophy of Science**, v. 63, p. 399-427, 2011.

LEVY, A. Three Kinds of New Mechanism. **Biology and Philosophy**, v. 28, p. 99-114, 2013.

LEVY, A. What was Hodgkin and Huxley's Achievement? **British Journal for the Philosophy of Science**, v. 65, p. 469-492, 2014.

LEVY, A.; BECHTEL, W. Abstraction and the Organization of Mechanisms. **Philosophy of Science**, v. 80, p. 241-261, 2012.

LITTLE, D. **Varieties of Social Explanation**: An Introduction to the Philosophy of Social Science. Boulder: Westview, 1991.

LITTLE, D. **Microfoundations, Method, and Causation**: On the Philosophy of the Social Science. New Brunswick: Transaction, 1998.

LONGINO, H. **Studying Human Behavior**: How Scientists Investigate Aggression and Sexuality. Chicago: University of Chicago Press, 2013.

LONGINO, H. Pluralism, Social Action, and the Causal Space of Human Behavior. **Metascience**, v. 23, p. 443-459, 2014.

MACHAMER, P. K. Teleology and Selective Processes. *In:* COLODNY, R. (ed.). **Logic, Laws, and Life**: Some Philosophical Complications. Pittsburgh: University of Pittsburgh Press, 1977, p. 129-142.

MACHAMER, P. Activities and Causation: The Metaphysics and Epistemology of Mechanisms. **International Studies in the Philosophy of Science**, v. 18, p. 27-39, 2004.

MACHAMER, P. K.; DARDEN, L.; CRAVER, C. F. Thinking about Mechanisms. **Philosophy of Science**, v. 67, p.1-25, 2000.

MACHERY, E. **Doing Without Concepts.** New York: Oxford University Press, 2009.

MACKIE, J. L. **The Cement of the Universe**. Oxford: Clarendon Press, 1974.

MALEY, C. J.; PICCININI, G. A Unified Mechanistic Account of Teleological Functions for Psychology and Neuroscience. *In:* KAPLAN, D. (ed.). **Integrating Mind and Brain Science**: Mechanistic Perspectives and Beyond. Oxford: Oxford University Press, 2017.

MATTHEWS, L. Embedded Mechanisms and Phylogenetics. **Philosophy of Science**, v. 85, n. 5, p. 1116, 2015. DOI: .

MATTHEWSON, J.; CALCOTT, B. Mechanistic Models of Population-Level Phenomena. **Biology and Philosophy**, v. 26, p. 737-756, 2011.

MCCAULEY, R. N.; BECHTEL, W. Explanatory Pluralism and the Heuristic Identity Theory. **Theory and Psychology**, v. 11, p. 736-760, 2001.

MELNYK, A. **A Physicalist Manifesto**: Thoroughly Modern Materialism, Cambridge: Cambridge University Press, 2003.

MELNYK, A. Comments on Sydney Shoemaker's Physical Realization. **Philosophical Studies**, v. 148, p. 113-123, 2010.

MENZIES, P. The Causal Structure of Mechanisms. **Studies in History and Philosophy of Biological and Biomedical Sciences**, v. 43, p. 796-805, 2012.

MILKOWSKI, M. **Explaining the Computational Mind.** Cambridge: MIT Press, 2013.

MILLSTEIN, R.L. Natural Selection as a Population-Level Causal Process. **The British Journal for the Philosophy of Science**, v. 57, p. 627-653, 2006.

MILO, R.; SHEN-ORR, S.; ITZKOVITZ, S.; KASHTAN, N.; CHKLOVSKII, D.; ALON, U. **Network Motifs**: Simple Building Blocks of Complex Networks. **Science**, v. 298, n. 5594, p. 824-827, 2002.

MITCHELL, S. D. Pragmatic Laws. **Philosophy of Science**, v. 64, p. S468-S479, 1997.

MITCHELL, S. D. Dimensions of Scientific Law. **Philosophy of Science**, v. 67, p. 242-265, 2000.

MITCHELL, S. D. **Biological Complexity and Integrative Pluralism.** Cambridge: Cambridge University Press, 2003.

MITCHELL, S. D. Modularity: More than a Buzzword? Essay Review. **Biological Theory**, v. 1, p. 98-101, 2005.

MITCHELL, S. D. **Unsimple Truths**: Science, Complexity, and Policy. Chicago: University of Chicago Press, 2009.

MOSS, L. Is the Philosophy of Mechanism Philosophy Enough? **Studies in the History and Philosophy of Science**: C, v. 43, p. 164-172, 2012.

NAGEL, E. **The Structure of Science**: Problems in the Logic of Scientific Explanation. New York: Harcourt, Brace and World, 1961.

NEANDER, K. The Teleological Notion of Function. **Australasian Journal of Philosophy**, v. 69, p. 454-468, 1991a.

NEANDER, K. Functions as Selected Effects: the Conceptual Analyst's Defence. **Philosophy of Science**, v. 58, p. 168-184, 1991b.

NICHOLSON, D. J. The Concept of Mechanism in Biology. **Studies in History and Philosophy of Biological and Biomedical Sciences**, v. 43, p. 152-163, 2012.

NICKELS, T. Beyond Divorce: Current Status of the Discovery Debate. **Philosophy of Science**, v. 52, p. 177-206, 1985.

OPPENHEIM, P.; PUTNAM, H. The Unity of Science as a Working Hypothesis. *In:* FEIGL, H.; SCRIVEN, M.; MAXWELL, G. (ed.). **Concepts, Theories and the Mind-Body Problem** (Minnesota Studies in the Philosophy of Science, v. 2). Minneapolis: University of Minnesota Press, 1958, p. 3-36.

PEARL, J. **Causality**: Models, Reasoning, and Inference, second edition. New York: Cambridge University Press, 2009.

PERSSON, J. Activity-Based Accounts of Mechanism and the Threat of Polygenic Effects. **Erkenntnis**, v. 72, p.135-149, 2010.

PICCININI, G. Computing Mechanisms. **Philosophy of Science**, v. 74, p. 501-526, 2007.

PICCININI, G.; Craver, C. F. Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches. **Synthese**, v. 183, p. 283-293, 2011.

PICCININI, G.; SCARANTINO, A. Information Processing, Computation, and Cognition. **Journal of Biological Physics**, v. 37, p. 1-38, 2011.

POLGER, T. W. Realization and the Metaphysics of Mind. **Australasian Journal of Philosophy**, v. 85, p. 233-259, 2007.

POLGER, T. W. Mechanisms and Explanatory Realization Relations. **Synthese**, v. 177, p. 193-212, 2010.

POPPER, K. **The Logic of Scientific Discovery**. London: Hutchinson and Co, 1959.

POVICH, M. Mechanisms and Model-Based Functional Magnetic Resonance Imaging. **Philosophy of Science**, v. 82, n. 5, p. 1035–1046, 2015. DOI: doi.org/10.1086/683438.

PSILLOS, S. A Glimpse of the Secret Connexion: Harmonizing Mechanism with Counterfactuals. **Perspectives on Science**, v. 12, p. 288-319, 2004.

RAMSEY, J. L. Mechanisms and Their Explanatory Challenges in Organic Chemistry. **Philosophy of Science**, v. 75, p. 970-982, 2008.

REICHENBACH, H. **Experience and Prediction**: An Analysis of the Foundations and the Structure of Knowledge. Chicago: The University of Chicago Press, 1938.

RICHARDSON, R. C.; STEPHAN, A. Emergence. **Biological Theory,** v. 2, p. 91-96, 2007.

ROE, S. **The Salmon-Roe Approach to Mechanistic Explanations.** Doctoral Dissertation, UC Davis, 2014.

ROMERO, F. Why there isn't interlevel causation in mechanisms. **Synthese,** v. 192, p. 3731-3755, 2015. DOI:10.1007/s11229-015-0718-0.

ROSENBERG, A. **The Structure of Biological Science**. Cambridge: Cambridge University Press, 1985.

RUSSELL, B. On the Notion of Cause. **Proceedings of the Aristotelian Society**, v. 13, p. 1-26, 1913.

RUSSO, F. **Causality and Causal Modeling in the Social Sciences**: Measuring Variation. Dordrecht: Springer, 2009.

RUSSO, F.; WILLIAMSON, J. Interpreting Causality in the Health Sciences. **International Studies in the Philosophy of Science**, v. 21, p. 157-170, 2007.

SALMON, W. C. **Scientific Explanation and the Causal Structure of the World.** Princeton: Princeton University Press, 1984.

SALMON, W. C. **Four Decades of Scientific Explanation.** Minneapolis: University of Minnesota Press, 1989.

SALMON, W. C. Causality Without Counterfactuals. **Philosophy of Science**, v. 61, p. 297-312, 1994.

SALMON, W. C. Causality and Explanation: A Reply to Two Critiques. **Philosophy of Science**, v. 64, p. 461-477, 1997.

SCHAFFER, J. Causation by Disconnection. **Philosophy of Science**, v. 67, p. 285-300, 2000.

SCHAFFER, J. Causes Need Not be Physically Connected to Their Effects: The Case for Negative Causation. *In:* HITCHCOCK, C. (ed.). **Contemporary Debates in Philosophy of Science.** Malden: Blackwell Publishing. 2004, p. 197-216.

SCHAFFNER, K. **Discovery and Explanation in Biology and Medicine**. Chicago: University of Chicago Press, 1993.

SCHLOSSER, G.; WAGNER, G. P. **Modularity in Development and Evolution**. Chicago: University of Chicago Press, 2004.

SCRIVEN, M. Explanation and Prediction in Evolutionary Theory. **Science**, v. 130, p. 477-482, 1959.

SHAGRIR, O. Computation, San Diego Style. **Philosophy of Science**, v. 77, p. 862-874, 2010.

SHAPIRO, L. Multiple Realizations. **Journal of Philosophy**, v. 97, p. 635-654, 2000.

SHOEMAKER, S. Realization, Micro-Realization, and Coincidence. **Philosophy and Phenomenological Research**, v. 67, p. 1-23, 2003.

SHOEMAKER, S. **Physical Realization**. Oxford: Oxford University Press, 1996.

SKIPPER Jr., R. A.; R. L. MILLSTEIN. Thinking about Evolutionary Mechanisms: Natural Selection. *In:* CRAVER, C. F.; DARDEN, L. **In Search of Mechanisms**: Discoveries Across the Life Sciences. Chicago: University of Chicago Press, 2013. p. 327-347.

SPIRTES, P.; GLYMOUR, C.; SCHEINES, R. **Causation, Prediction, and Search.** 2. ed. Cambridge: MIT Press, 2000.

STEEL, D.P. **Across the Boundaries**: Extrapolation in Biology and Social Science, New York: Oxford University Press, 2008.

STREVENS, M. **Depth**: An Account of Scientific Explanation. Cambridge: Harvard University Press, 2008.

SULLIVAN, J. The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-Reductionist Models of the Unity of Neuroscience, **Synthese**, v. 167, p. 511-539, 2009.

TABERY, J. Synthesizing Activities and Interactions in the Concept of a Mechanism. **Philosophy of Science**, v. 71, p. 1-15, 2004.

TABERY, J. Difference Mechanisms: Explaining Variation with Mechanisms. **Biology and Philosophy**, v. 24, p. 645-664, 2009.

TABERY, J. **Beyond Versus**: The Struggle to Understand the Interaction of Nature and Nurture. Cambridge: The MIT Press, 2014a.

TABERY, J. Pluralism, Social Action, and the Causal Space of Human Behavior. **Metascience**, v. 23, p. 443-459, 2014b.

TELLER, P. Mechanism, Reduction, and Emergence in Two Stories of the Human Epistemic Enterprise. **Erkenntnis**, v. 73, p. 413-425, 2010.

THAGARD, P. **Explaining Disease**: Causes, Correlations, and Mechanisms. Minds and Machines, v. 8, p. 61-78, 1998.

THAGARD, P. **How Scientists Explain Disease.** Princeton: Princeton University Press, 2000.

THAGARD, P. **Hot Thought:** Mechanisms and Applications of Emotional Cognition. Cambridge: The MIT Press, 2006.

THALOS, M. **Without Hierarchy**: An Essay on the Scale Freedom of the Universe. Oxford: Oxford University Press, 2013.

THEURER, K. L. Compositional Explanatory Relations and Mechanistic Reduction. **Minds and Machines**, v. 23, p. 287-307, 2013.

WEBER, M. **Philosophy of Experimental Biology**. Cambridge: Cambridge University Press, 2005.

WEBER, M. Causes without Mechanisms: Experimental Regularities, Physical Laws, and Neuroscientific Explanation. **Philosophy of Science**, v. 75, p. 995-1007, 2008.

WEISBERG, M. Three Kinds of Idealization. **The Journal of Philosophy**, v. 104, p. 639-659, 2007.

WEISBERG, M. **Simulation and Similarity**: Using Models to Understand the World. Oxford: Oxford University Press, 2013.

WEISKOPF, D. Models and Mechanisms in Psychological Explanation. **Synthese**, v. 183, p. 313-338, 2011.

WESTFALL, R. **The Construction of Modern Science.** Cambridge: Cambridge University Press, 1971.

WESTFALL, R. **Genes and the Agents of Life.** Cambridge: Cambridge University Press, 2005.

WIMSATT, W. C. Complexity and Organization. *In:* Kenneth F. SCHAFFNER, K. F.; COHEN, R. S. (ed.). **PSA 1972, Proceedings of the Philosophy of Science Association.** Dordrecht: Reidel, 1972a, p. 67-86.

WIMSATT, W. C. Teleology and the Logical Structure of Function Statements", **Studies in the History and Philosophy of Science**, v. 3, p.1-80, 1972b.

WIMSATT, W. C. Reductionism, Levels of Organization, and the Mind–Body Problem. *In:* GLOBUS, G.; SAVODNIK, I.; MAXWELL, G. (ed.). **Consciousness and the Brain.** New York: Plenum Press, 1976, p. 199-267.

WIMSATT, W. C. Aggregativity: Reductive Heuristics for Finding Emergence. **Philosophy of Science**, v. 64, p. S372-S384, 1997.

WIMSATT, W. C. **Re-Engineering Philosophy for Limited Beings**: Piecewise Approximations to Reality. Cambridge: Harvard University Press, 2007.

WOODWARD, J. The Causal/Mechanical Model of Explanation. *In:* KITCHER, P.; SALMON, W. C. (ed.). **Scientific Explanation**, (Minnesota Studies in the Philosophy of Science 13), p. 357-383, 1989.

WOODWARD, J. Law and Explanation in Biology: Invariance Is the Kind of Stability that Matters. **Philosophy of Science**, v. 68, p. 1-20, 2001.

WOODWARD, J. What Is a Mechanism?: A Counterfactual Account. **Philosophy of Science**, v. 69, p. S366-S377, 2002.

WOODWARD, J. **Making Things Happen**: A Theory of Causal Explanation. New York: Oxford University Press, 2003.

WOODWARD, J. Mechanisms Revisited. **Synthese**, v. 183, p. 409-427, 2011.

WOODWARD, J. A Functional Account of Causation; or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters -Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment). **Philosophy of Science**, v. 81, p. 691-713, 2014.

WRIGHT, L. Functions. **Philosophical Review**, v. 82, p. 139-168, 1973.

YLIKOSKI, P. Causal and Constitutive Explanation Compared. **Erkenntnis**, v. 78, p. 277-297, 2013.

(VIII) Simulações Computacionais na Ciência*

Autor: Eric Winsberg
Tradução: Nathália C. A. P. Strongen
Revisão: Felipe Rocha

A simulação computacional foi pioneira enquanto ferramenta científica nos estudos meteorológicos e de física nuclear no período imediatamente após a II Guerra Mundial e, desde então, tornou-se indispensável em um crescente número de disciplinas. A lista de ciências que fazem extensivo uso da simulação computacional tem crescido de modo a incluir astrofísica, física de partículas, ciência dos materiais, engenharia, mecânica dos fluidos, ciência climática, biologia evolucionária, ecologia, economia, teoria da decisão, medicina, sociologia, epidemiologia e muitas outras. Existem, ainda, algumas disciplinas, como a teoria do caos e teoria da complexidade, cuja própria existência emergiu a partir do desenvolvimento dos modelos computacionais que permeavam estudos anteriores.

---

* WINSBERG, E. Computer Simulations in Science. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2011. Disponível em: https://plato.stanford.edu/entries/simulations-science/. Acesso em: 22 set. 2021.

The following is the translation of the entry on Computer Simulations in Science by Eric Winsberg, 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/simulations-science/. 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/simulations-science/. 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.

Após um início lento, filósofos da ciência começaram a dedicar mais atenção ao papel da simulação na ciência. Várias questões de interesse filosófico em relação à simulação computacional têm emergido, tais como: Qual é a estrutura epistemológica da simulação computacional? Qual é a relação entre simulação computacional e experimento? A simulação computacional gera questões para a filosofia da ciência que não são totalmente abarcadas pelo trabalho recente sobre modelos em geral? O que a simulação computacional nos ensina sobre emergência? Sobre a estrutura das teorias científicas? Qual é o papel (se existe) das ficções na modelagem científica?

1. O que é Simulação Computacional?

Nenhuma definição única de simulação computacional é apropriada. Primeiro, o termo é usado tanto em um sentido estrito quanto em um sentido amplo. Em segundo lugar, pode-se desejar entender o termo sob mais de um ponto de vista.

1.1 Uma definição estrita.

Em um sentido estrito, uma simulação computacional é um programa que funciona em um computador a partir de métodos passo a passo a fim de explorar o comportamento aproximado de um modelo matemático.

Usualmente, tal modelo diz respeito a um sistema do mundo real (que também pode ser imaginário ou hipotético). Nesse sentido, um programa de computador é um **modelo de simulação computacional**. Assim, uma execução do programa no computador é uma simulação computacional do sistema. O algoritmo considera como entrada uma especificação do estado do sistema (o valor de todas as variáveis) em um tempo **t**. Então é calculado o estado do sistema no tempo **t**+1. A partir dos valores que caracterizam o segundo estado, é calculado o estado do sistema no tempo **t**+2, e assim por diante. Ao ser executado em um computador, o algoritmo produz uma imagem numérica do desenvolvimento dos estados do sistema, tal como é conceitualizado no modelo.

Esta sequência de valores variáveis para o modelo pode ser salva em uma grande coleção de “informações” e ser frequentemente visualizada na tela do computador utilizando métodos específicos para tal ação. Frequentemente, mas certamente não sempre, os métodos de visualização são desenvolvidos para **imitar**

o resultado equivalente ao de algum instrumento científico, de modo que a simulação possa ser mensurada como um sistema relevante.

Às vezes, os métodos passo a passo da simulação computacional são utilizados porque o modelo de interesse contém equações (diferenciais) contínuas (as quais especificam continuamente o grau de mudança no tempo) que não podem ser resolvidas analiticamente - em princípio ou talvez apenas em prática. Essa perspectiva segue na mesma esteira que a definição dada por Paul Humphreys: "qualquer método computacional é implementado para exploração das propriedades dos modelos matemáticos nos quais métodos analíticos não são viáveis" (1991, p. 500). Mas até mesmo enquanto definição estrita, esta deve ser lida cuidadosamente e não tomada como sugerindo que simulações são apenas utilizadas quando se trata de equações que não são resolvidas analiticamente no próprio modelo. Simulações computacionais são frequentemente utilizadas tanto quando o modelo original contém equações discretas - e quando o modelo original consiste em algo melhor descrito como **regras de desenvolvimento** do que **equações**.

No primeiro caso, quando equações são "discretizadas" (tornando as equações que descrevem variações contínuas em equações discretas), deve ser enfatizado que, embora seja comum falarmos que as simulações "resolvem" tais equações, a discretização pode no melhor dos casos possíveis apenas oferecer um resultado que se aproxima da solução das equações contínuas, com algum grau de precisão. Finalmente, quando falamos de uma "simulação computacional" em um sentido estrito, nós devemos falar de uma implementação particular de um algoritmo em um computador digital particular, desenvolvido em uma linguagem particular, usando um compilador particular, etc. Há casos em que diferentes resultados podem ser obtidos a partir da variação de qualquer um desses particulares.

1.2 Uma definição geral

De forma mais geral, podemos pensar as simulações computacionais como um método explicativo para o estudo de sistemas. Nesta definição geral do termo, isso se refere ao processo como um todo. Tal processo inclui a escolha de um modelo; uma maneira de implementá-lo de uma forma que possa funcionar em um computador; o cálculo dos resultados do algoritmo; e a visualização e o estudo da informação resultante. O método inclui o processo completo, utilizado para realizar inferências sobre o sistema final que se pretende modelar, assim como os processos

necessários para sancionar estas inferências. Esta concepção se assemelha à definição de simulação computacional oferecida por Winsberg (2003, p. 111):

> Estudos de sucesso sobre simulação oferecem mais do que calcular números. Neles são utilizadas técnicas variáveis para extrair inferências desses números. Nesse sentido, simulações realizam um uso criativo das técnicas de cálculo que podem apenas ser fundamentadas de forma extra matemática ou extra teórica. Como tal, ao contrário de cálculos simples que podem ser realizados em um computador, os resultados das simulações não são automaticamente confiáveis. Muito esforço e conhecimento estão envolvidos na decisão de quais resultados da simulação são confiáveis e quais não são.

Quando filósofos da ciência escrevem sobre simulação computacional e argumentam sobre quais propriedades epistemológicas ou metodológicas as "simulações computacionais" possuem, eles geralmente querem dizer sobre o entendimento do termo em um sentido amplo de estudos computacionais.

1.3 Um ponto de vista alternativo

As duas definições acima consideram a simulação computacional sendo fundamentalmente sobre o uso de um computador para resolver, ou para resolver aproximadamente, as equações matemáticas de um modelo que pretende representar algum sistema - real ou hipotético. Outra abordagem tenta definir "simulação" independentemente da noção de simulação computacional e, em seguida, definir "simulação computacional" composicionalmente: como uma simulação realizada por um computador digital programado. Nesta abordagem, uma simulação é qualquer sistema em que se acredita, ou espera, ter um comportamento dinâmico semelhante o suficiente a algum outro sistema, de modo que o primeiro possa ser estudado para se inferir algo sobre o último.

Por exemplo, se estudarmos algum objeto, porque acreditamos que seja suficientemente e dinamicamente semelhante a uma bacia de fluído para que possamos aprender algo sobre elas analisando tal objeto, então uma simulação de

bacias de fluído é gerada. Isso está de acordo com a definição de simulação que encontramos em Hartmann: algo que "imita um processo por outro processo. Nesta definição, o termo 'processo' refere-se apenas a algum objeto ou sistema no qual o estado se altera no tempo" (1996, p. 83). Hughes (1999) objetou que a definição de Hartmann descarta simulações que imitam a estrutura de um sistema e não sua dinâmica. Humphreys revisou sua definição de simulação de acordo com as observações de Hartmann e Hughes, como se segue:

> O Sistema **S** fornece uma simulação central de um objeto ou processo **B** apenas no caso de **S** ser um dispositivo computacional concreto que produz, através de um processo temporal, soluções para um modelo computacional [...] que representa corretamente **B**, tanto dinâmica quanto estaticamente. Se, além disso, o modelo computacional usado por **S** representa corretamente a estrutura do sistema real **R**, então **S** fornece uma simulação central do sistema **R** em relação ao **B** (2004, p. 110).

(Observe que Humphreys está aqui definindo simulações computacionais, e não simulações em geral. Humphreys está fazendo isso com o objetivo de definir um termo da composição). Deve-se notar que as definições de Humphreys impedem o termo simulação de ser um sucesso, o que é lamentável. Uma definição melhor seria aquela que, como será apresentado na última seção, inclui termos como "acredita" ou "espera" a fim de tentar oferecer uma definição satisfatória.

Na maioria das discussões filosóficas sobre simulações computacionais, o conceito mais útil é o definido na seção 1.2. A exceção é quando o objetivo da discussão é explicitamente entender a simulação computacional como um exemplo de simulação geral (*vide* seção 5). Exemplos de simulações que não são computacionais incluem o famoso modelo físico da Baía de São Francisco (HUGGINS; SCHULTZ, 1973). Este é um modelo em escala hidráulica do funcionamento do sistema da baía de São Francisco e do delta do rio Sacramento - *San Joaquin*, construídos na década de 1950 pelo Corpo de Engenheiros do Exército a fim de analisar possíveis intervenções de engenharia na Baía. Outro bom exemplo, discutido extensivamente em (DARDASHTI *et al*., 2015, 2019), é o uso de "Buracos burros" acústicos formados de condensados de Bose-Einstein para estudar o comportamento dos buracos negros. O físico Bill Unruh observou que, em contextos dinâmicos, algo semelhante

a um buraco negro poderia surgir se houvessem regiões ativas que se movimentassem rápido o bastante, as ondas teriam que se mover mais rápido que a velocidade do som (algo que elas não podem fazer) (UNRUH, 1981). Tais regiões teriam, de fato, eventos sônicos horizontais. Unruh chamou essa configuração física de "buraco burro" ("burro" no sentido de "sem sentido") e propôs que este fenômeno pudesse ser estudado para aprender aspectos que ainda não conhecemos acerca dos buracos negros. Por algum tempo, essa proposta foi vista como nada mais do que uma idéia inteligente, mas recentemente os físicos perceberam que, usando os condensados de Bose-Einstein, é realmente possível construir e estudar buracos burros em laboratório. Está claro por que devemos pensar nessa configuração como uma simulação: o buraco burro simula o buraco negro. Em vez de encontrar um programa de computador para simular os buracos negros, os físicos encontraram uma configuração dinâmica fluida a qual se acredita ser um bom modelo e para o qual esse modelo tem semelhanças matemáticas fundamentais com o modelo dos sistemas de interesse. Os físicos observam o comportamento da organização dos fluídos em laboratório para realizar inferências sobre os buracos negros. Assim, as definições de simulação nesta seção focam na tentativa de compreender em qual sentido a simulação computacional e esse tipo de atividade são atividades do mesmo tipo. Podemos estar em uma situação melhor para entender por que uma simulação no sentido de 1.3 que é executada em um computador se sobrepõe a uma simulação no sentido de 1.2. Voltaremos a isso na seção 5.

Barberousse *et al*. (2009), no entanto, criticam essa analogia. Os autores ressaltam que as simulações computacionais não funcionam da mesma maneira que a simulação de Unruh. Não é o caso que o computador como objeto material e o sistema de destino sigam as mesmas equações diferenciais. Uma boa referência sobre simulações que não são simulações computacionais pode ser vista em Trenholme (1994).

2. Tipos de simulações computacionais

Dois tipos de simulação computacional são frequentemente distinguidos: **simulações baseadas em equações** e **simulações baseadas em agentes** (ou baseada em indivíduos). Ambos os tipos de simulações computacionais são utilizados para três finalidades distintas: previsão (pontual e global/qualitativa), compreensão e exploração ou propósito heurístico.

2.1 Simulações baseadas em equações

Simulações baseadas em equações são comumente usadas nas ciências físicas e em outras ciências nas quais existe uma teoria central que orienta a construção de modelos matemáticos com base em equações diferenciais. Aqui é usado o termo “baseado em equações” para se referir a simulações baseadas nos tipos de equações globais que são associadas a teorias físicas, em oposição a "regras da evolução" (que serão discutidas na próxima seção). As simulações baseadas em equações podem também ser baseadas em particulares, onde existem muitos corpos discretos e um conjunto de equações diferenciais que governam a interação, ou as simulações podem ser baseadas em campo, onde existe um conjunto de equações que governam a evolução temporal de um meio ou campo contínuo. Um exemplo da primeira situação citada é a simulação da formação das galáxias, na qual a interação gravitacional entre uma coleção finita de corpos discretos é discretizada no tempo e no espaço. Um exemplo da última situação é a simulação de uma situação dinâmica, como um sistema meteorológico do tipo de uma tempestade severa. Aqui, o sistema é tratado como um meio contínuo - dinâmico - e um campo que representa a distribuição das variáveis relevantes no espaço é discretizado no espaço e, em seguida, atualizado em intervalos discretos de tempo.

2.2 Simulações baseadas em agentes

Simulações baseadas em agentes são mais comuns nas ciências sociais e comportamentais, embora também as encontremos em disciplinas como vida artificial, epidemiologia, ecologia e qualquer disciplina em que a rede interação de muitos indivíduos está sendo estudada. As simulações baseadas em agentes são semelhantes as que são construídas a partir de particulares nas quais está sendo representado o comportamento de **n** indivíduos distintos. Porém, diferentemente das simulações baseadas em partes de equações, não existem equações diferenciais globais que governam os movimentos dos indivíduos. Ao contrário, em simulações baseadas em agentes, o comportamento dos indivíduos é ditado por suas próprias regras locais.

Para ilustrar: uma simulação famosa e inovadora baseada em agentes foi a de Thomas Schelling (1971), o modelo de “segregação”. Os agentes em sua simulação eram indivíduos que "viviam" em um tabuleiro de xadrez. Os indivíduos foram divididos em dois grupos característicos (por exemplo, duas raças diferentes, meninos e meninas, fumantes e não fumantes etc.). Cada quadrado do tabuleiro representava uma casa, com no máximo uma pessoa. Um indivíduo ficaria feliz se tivesse certa porcentagem de vizinhos de seu próprio grupo. Agentes felizes ficam onde estão e agentes infelizes se mudam para locais que estão disponíveis. Schelling descobriu que a configuração do tabuleiro evoluiu rapidamente para um padrão de localização fortemente segregado se as "regras de felicidade" dos agentes fossem especificadas para que tal segregação fosse favorecida. Surpreendentemente, no entanto, ele também descobriu que as partes do tabuleiro inicialmente integradas estavam totalmente segregadas, mesmo que as regras de felicidade dos agentes expressassem apenas uma leve preferência por ter vizinhos de seu próprio tipo.

2.3 Simulações de múltiplas escalas

Na seção 2.1, discutimos modelos baseados em equações que são também baseados em métodos particulares e aqueles que são baseados em métodos de campo. Mas alguns modelos de simulação são híbridos e, portanto, constituídos de diferentes tipos de métodos de modelagem. Os modelos de simulação em múltiplas escalas, em particular, unem elementos de modelagem de diferentes escalas de descrição. Um bom exemplo disso seria um modelo que simula a dinâmica da massa material tratando a matéria como um campo submetido a tensões e deformações em um nível de descrição relativamente grosseiro, mas que focaliza regiões específicas do material onde estão ocorrendo importantes efeitos em pequena escala, e modela essas regiões menores com métodos de modelagem refinados. Tais métodos podem depender da dinâmica molecular, ou mecânica quântica, ou ambos - cada um destes é uma descrição mais detalhada da matéria do que as oferecidas tratando o material como um campo. Os métodos de simulação em múltiplas escalas podem ser divididos em métodos multiescalas seriais e multiescalas paralelas. O método mais tradicional é a modelagem multiescala serial. A idéia aqui é escolher uma região, simulá-la no nível mais baixo da descrição, resumir os resultados em um conjunto de parâmetros compreendidos pelo modelo de nível superior e passá-los para a parte do algoritmo que funciona em um nível superior.

Os métodos de modelagem multiescala serial não são eficazes quando as diferentes escalas são fortemente acopladas. Quando escalas diferentes interagem para produzir o comportamento observado, neste caso é necessária uma abordagem que simule cada região simultaneamente. Esse método é chamado de modelagem multiescala paralela. A modelagem em múltipla escala paralela é a base de um método de simulação quase ubíqua: a chamada modelagem de "subvalor". A modelagem de subvalor refere-se à representação de importantes processos físicos de pequena escala que ocorrem de maneira que não podem ser adequadamente resolvidos a partir da dimensão dos valores de uma simulação específica (lembre-se de que muitas simulações discretizam equações contínuas, possuindo, assim, uma "sequência de valores" potencialmente finita). Nos estudos acerca da turbulência em sistemas dinâmicos, por exemplo, uma estratégia prática comum para realizar certos cálculos é levar em conta os **vórtices** de pequena escala (ou **espirais**) que são abarcados pela rede global. Isso é feito adicionando ao movimento de larga escala uma **viscosidade turbulenta** que caracteriza o transporte e a dissipação de energia no fluxo de menor escala, ou qualquer recurso que ocorra em uma escala muito pequena para ser capturado pela rede global.

Nas ciências climáticas e nas disciplinas afins, a modelagem de subvalores é chamada de "parametrização". Isso, novamente, refere-se ao método de substituição de processos, aqueles que são muito pequenos ou complexos demais para serem representados fisicamente no modelo, por uma descrição matemática mais simples. Isso é diferente de outros processos, como por exemplo, o fluxo em larga escala da atmosfera, calculado a partir de valores de acordo com a teoria básica. A modelagem é chamada de "parametrização" porque vários **parâmetros** não físicos são necessários para conduzir os algoritmos que calculam os subvalores. Exemplos de parametrização na simulação climática incluem a taxa descendente de gotas de chuva, a taxa de transferência radiativa atmosférica e a taxa de formação de nuvens. Por exemplo, a nebulosidade média acima de 100 km^2 não está claramente relacionada à umidade média considerando-se o parâmetro da simulação. No entanto, à medida que a umidade média aumenta, a nebulosidade média também aumenta, portanto, pode haver um parâmetro que vincule a nebulosidade média à umidade média em algum contexto. Embora as parametrizações modernas da formação de nuvens sejam mais sofisticadas do que isso, a ideia básica é bem ilustrada pelo exemplo. O uso de métodos de modelagem de subvalor em simulações acarreta consequências importantes para a compreensão da estrutura da epistemologia da simulação. Isto será discutido em mais detalhes na seção 4.

Os métodos de modelagem de subvalores podem ser contrastados com modelos multiescala em paralelo, em que os algoritmos de subvalores são teoricamente baseados em princípios, mas são motivados por uma teoria em um nível diferente de descrição. No exemplo da simulação da massa material mencionada acima, o algoritmo que conduz o menor nível de descrição não é construído pela própria experiência. O algoritmo que conduz o nível menor é, na verdade, teoricamente mais baseado em princípios do que o nível superior, no sentido de que a física é mais fundamental: mecânica quântica ou dinâmica molecular *versus* mecânica contínua. Esses tipos de modelos de múltiplas escalas, em outras palavras, unem recursos de ambas as teorias em diferentes níveis de descrição. Tais modelos fornecem exemplos interessantes que motivam a reflexão acerca das relações inter teóricas, que desafiam a visão amplamente aceita de que um conjunto de leis inconsistentes não pode ter modelos.

2.4 Simulações de Monte Carlo

Na literatura científica, há outra grande classe de simulações computacionais chamada de Simulações de Monte Carlo (MC). Simulações MC são algoritmos computacionais que utilizam a aleatoriedade para calcular as propriedades de um modelo matemático nos quais algoritmos randômicos não são característicos do modelo final. Um bom exemplo é o uso de um algoritmo aleatório para calcular o valor de π. Se desenharmos uma unidade quadrada em um pedaço de papel e inscrever um círculo sob ele, e depois inserir aleatoriamente uma coleção de objetos sob o quadrado, a proporção de objetos que pousam no círculo seria aproximadamente igual a π/4. Uma simulação de computador que simula um procedimento como esse seria chamada de simulação MC para calcular π.

Muitos filósofos da ciência se desviaram da linguagem científica comum e se esquivaram de pensar em simulações MC como simulações genuínas. Grüne-Yanoff e Weirich (2010) oferecem o raciocínio a seguir: “A abordagem de Monte Carlo não tem um propósito mimético: imita um sistema determinístico não para servir como seu substituto na investigação, mas apenas para oferecer um cálculo alternativo das propriedades do sistema determinístico" (p. 30). Isso mostra que as simulações MC não se encaixam em nenhuma das definições acima. Por outro lado, a divisão entre os filósofos e a linguagem comum pode ser ajustada ao observar que as simulações MC simulam um processo imaginário que pode ser usado para

calcular algo relevante para analisar outro processo. Suponha que uma órbita planetária esteja sendo modelada e, para o cálculo, precisamos saber o valor de π. Se realizarmos a simulação MC mencionada no último parágrafo, estamos simulando o processo de soltar objetos aleatoriamente em um espaço, mas o que estamos modelando é uma órbita planetária. É nesse sentido que as simulações MC são simulações, mas não são simulações dos sistemas básicos do estudo em questão. No entanto, como Beisbart e Norton (2012) apontam, algumas simulações MC (ou seja, aquelas que usam técnicas de MC para resolver equações dinâmicas estocásticas referentes a um sistema físico) são de fato simulações dos sistemas que está sendo analisado.

3. Propósitos da simulação

Existem três finalidades gerais para as quais simulações computacionais podem ser aplicadas. Uma delas é que as simulações podem ser utilizadas para fins heurísticos, com o objetivo de prever dados que não possuímos e para entender aqueles que temos acesso.

Na categoria de modelos heurísticos, as simulações podem ser subdivididas em aquelas utilizadas para divulgar conhecimento a outras pessoas e àquelas utilizadas para representar informações para nós mesmos. Quando Watson e Crick brincava com chapas e arames, eles estavam fazendo o último no começo e o primeiro quando mostraram os resultados para os outros. Quando o Exército construiu o modelo da Baía de São Francisco para convencer a população a votar a favor de uma intervenção em particular que era considerada perigosa, utilizaram as simulações heurísticas com o propósito do primeiro tipo. As simulações computacionais podem ser usadas para esses dois tipos de propósitos - para explorar recursos de possíveis estruturas representacionais; ou comunicar conhecimento a outras pessoas. Por exemplo: simulações computacionais de processos naturais, como reprodução bacteriana, deslocamento tectônico, reações químicas e evolução, nesses casos as simulações foram até utilizadas em salas de aulas para ajudar os alunos a visualizar a estrutura interior de fenômenos e processos que são impraticáveis, impossíveis ou dispendiosos para ilustrar em um ambiente laboratorial “de fato”.

Outra classe ampla de propósitos para os quais simulações computacionais podem ser utilizadas é oferecer uma perspectiva acerca do que se esperar em relação ao comportamento de um sistema no mundo real sob um conjunto específico

de circunstâncias. Resumidamente falando: a simulação computacional pode ser usada para previsão. Podemos utilizar modelos para prever o futuro, ou para inferir algo sobre o passado; podemos utilizá-los para fazer previsões precisas ou gerais. Nesse momento, detalhamos nossa taxonomia no que diz respeito à precisão das previsões que podemos realizar a partir de simulações. Existem a) previsões pontuais: onde estará o planeta Marte em 21 de outubro de 2300? b) Previsões "qualitativas", globais ou sistêmicas: a órbita deste planeta é estável? Que lei de escala emerge desses tipos de sistemas? Qual é a dimensão fractal do atrator para sistemas desse tipo? e C) Previsões de alcance: é 66% provável que a temperatura média global da superfície aumente entre 2 e 5 graus centígrados até o ano 2100; é "altamente provável" que o nível do mar suba em pelo menos dois pés; é "implausível" que a circulação termoalina desapareça nos próximos 50 anos.

Finalmente, simulações podem ser usadas para entender os sistemas e seus comportamentos. Se já temos dados sobre o comportamento do sistema, podemos usar a simulação computacional para responder a perguntas sobre como esses eventos poderiam ter ocorrido; ou sobre como esses eventos realmente ocorreram.

Ao pensar no tópico da próxima seção, a epistemologia das simulações computacionais, devemos também ter em mente que os procedimentos necessários para sancionar os resultados das simulações dependerão, frequentemente e em grande parte, de qual do tipo ou para qual finalidade acima será aplicada.

4. A Epistemologia das Simulações Computacionais

Como os métodos de simulação por computador ganharam importância em muitas disciplinas, o estudo acerca da confiabilidade para gerar novos conhecimentos aumentou, especialmente quando se espera que simulações sejam consideradas epistemicamente equivalentes a experimentos e métodos teóricos analíticos tradicionais. A questão presente é sempre se os resultados de uma simulação computacional específica são precisos o suficiente para sua finalidade. Se uma simulação estiver sendo usada para prever o clima, ela prevê as variáveis em que estamos interessados em um grau de precisão suficiente para atender às necessidades de seus consumidores? Se uma simulação da atmosfera acima de uma planície do centro-oeste está sendo utilizada para **entender** a estrutura de uma tempestade severa, temos confiança de que as estruturas envolvidas, aquelas que desempenharão um papel explicativo no entendimento das características do

fenômeno como o fato de que a tempestade às vezes se divide ou que, às vezes, formam tornados, estão sendo abarcadas com precisão suficiente para justificar a confiança na explicação? Se uma simulação abarcar elementos de engenharia e *design*, as previsões feitas pela simulação são confiáveis o suficiente para sancionar uma escolha específica de parâmetros de *design* ou para sancionar nossa crença de que um design específico da asa do avião funcionará? Assumindo que a resposta a essas perguntas às vezes é "sim", ou seja, que esses tipos de inferências são pelo menos às vezes justificadas, a questão filosófica central é: o que as justifica? De maneira mais geral, como se pode afirmar que uma simulação é boa o suficiente para a finalidade que foi desenvolvida? Estas são as questões centrais da epistemologia da simulação computacional (ESC).

Dado que a teoria da confirmação é um dos tópicos tradicionais da filosofia da ciência, pode parecer óbvio que este ramo teria recursos para começar a abordar essas questões. Winsberg (1999), no entanto, argumentou que, quando se trata de tópicos relacionados à necessidade do conhecimento, a filosofia da ciência tradicionalmente se preocupa com a justificativa de teorias, não com sua aplicação. A maioria das simulações, por outro lado, na medida em que se baseiam em teorias, tendem a utilizar as mais estabelecidas. A ESC, em outras palavras, raramente trata de testar as teorias básicas que podem ser utilizadas na simulação e, mais frequentemente, estabelece a credibilidade das hipóteses que são, em parte, o resultado das aplicações dessas teorias.

4.1 Novos recursos da ESC

Winsberg (2001) argumenta que, diferentemente das questões epistemológicas que ocupam papel central em relação às teorias da confirmação, uma ESC adequada deve atender a três condições. Em particular, é necessário que seja considerado o fato de que o conhecimento produzido por simulações computacionais é resultado de inferências **descendentes**, **heterogêneas** e **autônomas**.

Descendente. A ESC deve refletir o fato de que, em muitos casos, é a partir das teorias científicas aceitas que se inicia a construção de modelos de simulação computacional, desempenhando um papel importante no processo desde a inferência dos resultados da simulação até as conclusões sobre o sistema do mundo real. A palavra "descendente" foi utilizada no sentido de que, diferentemente das maiorias das inferências científicas que são interessantes aos filósofos, as que

partem *da* observação da instanciação de teorias, aqui temos inferências que são extraídas (em parte) de teorias gerais, **descendendo** até características particulares dos fenômenos.

Heterogêneo. A ESC deve considerar que os resultados das simulações, geralmente dependem não apenas da teoria, mas também de outros elementos e recursos do modelo, incluindo parametrizações (discutidas anteriormente), métodos de solução numérica, artifícios matemáticos, aproximações e idealizações, ficções, suposições *ad hoc*, conjunto de funções, compiladores e *hardware* computacional e, talvez o mais importante, sangue, suor e lágrimas de muitas tentativas e erros.

Autônomo. A ESC deve levar em consideração a autonomia do conhecimento produzido por meio de simulações na medida em que tal conhecimento não pode ser totalmente sancionado se comparado com a observação. Simulações são geralmente empregadas para estudar fenômenos nos quais os dados são escassos. Nestas circunstâncias, simulações têm como objetivo substituir experimentos e observações como fontes de dados sobre o mundo, pois tais ferramentas estão fora de alcance, por razões de princípios, práticas ou éticas.

Parker (2013) afirmou que a utilidade dessas condições é um pouco comprometida pelo fato de estar excessivamente focadas na simulação nas ciências físicas e em outras disciplinas orientadas por teorias e baseadas em equações. Isso parece correto. Nas ciências sociais e comportamentais e em outras disciplinas em que a simulação é baseada em agentes (*vide* 2.2) o foco deve ser nas normas e onde os modelos são construídos na ausência de teorias estabelecidas e quantitativas, nesse caso a ESC provavelmente deveria ser caracterizado em outros termos.

Por exemplo, alguns cientistas sociais que utilizam simulações baseadas em agentes seguem uma metodologia na qual os fenômenos sociais (por exemplo, um padrão observado como a segregação) são explicados, ou analisados, gerando fenômenos semelhantes às suas simulações (EPSTEIN; AXTELL, 1996; EPSTEIN, 1999). Mas esta metodologia em particular traz à tona várias questões epistemológicas. O que exatamente foi realizado, que tipo de conhecimento foi adquirido quando um fenômeno social observado é mais ou menos reproduzido por uma simulação baseada em agentes? Essa simulação pode ser considerada uma explicação do fenômeno? Uma possível explicação? (*vide* GRÜNE-YANOFF, 2007). Giuseppe Primiero (2019) argumenta que existe um domínio de "ciências artificiais" construído em torno de simulações baseadas em agentes e em sistemas multiagentes, que requer sua própria epistemologia - aquela em que a validação não pode ser definida por comparação com o mundo real, mas deve ser definida em relação ao sistema **pretendido**.

Também é justo dizer, como Parker faz (2013), que as condições descritas acima prestam pouca atenção aos diversos e diferentes propósitos para os quais as simulações são utilizadas (como discutido em 2.4). Se estamos utilizando uma simulação para fazer previsões quantitativas detalhadas sobre o comportamento futuro de um sistema específico, a epistemologia de tais inferências pode exigir padrões mais rigorosos do que aqueles envolvidos quando as inferências feitas são sobre o comportamento geral e qualitativo de uma classe de sistemas. De fato, também é justo dizer que poder-se-ia desprender mais esforços em relação à classificação dos tipos de propósitos para os quais as simulações computacionais são aplicadas e as limitações que tais propósitos impõem à estrutura da própria epistemologia.

Frigg e Reiss (2009) argumentaram que nenhuma dessas três condições é nova na simulação computacional. Os autores apontaram que a modelagem comum de "papel e lápis" incorpora esses recursos. Na verdade, defenderam que a simulação computacional não poderia levantar novas questões epistemológicas porque estas poderiam ser abarcadas na análise sobre a adequação do modelo subjacente à simulação, que é idêntico aos problemas epistemológicos que surgem na modelagem comum e à questão da correção da solução para as equações do modelo fornecidas pela simulação, que é uma questão matemática, e não relacionada à epistemologia da ciência. No primeiro ponto, Winsberg (2009b) respondeu que era a confluência simultânea dos três recursos que era nova na simulação. Voltaremos ao segundo ponto na seção 4.3.

4.2 ESC e a epistemologia do experimento

Alguns dos trabalhos de ESC desenvolveram analogias entre simulação computacional e experimentação, recorrendo a trabalhos recentes na epistemologia do experimento, particularmente o de Allan Franklin.

Em seu trabalho sobre a epistemologia do experimento, Franklin (1986, 1989) identificou uma série de estratégias que os experimentadores usam para aumentar a confiança racional em seus resultados. Weissart (1997) e Parker (2008a) defenderam várias formas de analogia entre essas estratégias e outras disponíveis para os simulacionistas para sancionar seus resultados. A análise mais detalhada desses trabalhos pode ser encontrada em Parker (2008a), onde o autor também usa essas analogias para destacar pontos fracos nas abordagens atuais da avaliação do modelo de simulação.

Winsberg (2003) também se baseia no trabalho de Ian Hacking (1983, 1988, 1992) sobre a filosofia do experimento. Um dos *insights* centrais de Hacking acerca do experimento está contido em seu *slogan* de que os experimentos têm vida própria (1992, p. 306). Hacking pretendia contemplar duas coisas com esse *slogan*. A primeira foi uma reação contra a imagem instável da ciência que vem, por exemplo, de Kuhn. Hacking (1992) sugere que os resultados experimentais podem permanecer estáveis mesmo diante de mudanças drásticas na ciência. O segundo ponto que Hacking pretendia transmitir era que "os experimentos são orgânicos, desenvolvem-se, mudam e ainda mantêm certo desenvolvimento a longo prazo que nos faz falar sobre repetir e replicar experimentos:" (1992, p. 307). Algumas das técnicas que os simulacionistas usam para construir seus modelos são credenciadas da mesma maneira que Hacking diz que instrumentos, procedimentos e métodos experimentais o fazem; as credenciais se desenvolvem por um longo período de tempo e tornam-se profundamente vinculadas à tradição. Segundo Hacking a linguagem, as técnicas e os conjuntos de suposições que os simulacionistas utilizam tornam-se "auto-justificativas". Talvez uma expressão melhor fosse que eles carregam suas próprias credenciais. Isso fornece uma resposta ao problema apresentado em 4.1, de entender como a simulação poderia ter uma epistemologia viável, apesar da natureza heterogênea e autônoma de suas inferências.

Inspirado por outro filósofo especialista em experimentos (MAYO, 1996), Parker (2008b) sugere uma tentativa de solução para algumas das limitações das abordagens atuais para a avaliação de modelos de simulação. Neste trabalho, Parker sugere que a abordagem estatística de Mayo para a análise de erros e o consequente entendimento dos experimentos tradicionais, utilizando a noção de "teste severo", poderia lançar luz sobre a epistemologia da simulação. A questão central da epistemologia da simulação, do ponto de vista estatístico do erro, consiste em: "O que justifica considerar uma simulação computacional um teste severo de algumas hipóteses sobre o mundo natural? Em outras palavras, o que garante nossa conclusão de que é improvável que a simulação ofereça os resultados que realmente foram disponibilizados, se a hipótese de interesse for falsa (2008b, p. 380)? Parker acredita que muito do que é necessário para avaliar o modelo de simulação carece de rigor e estrutura, pois:

> consiste em mais do que comparações lado a lado do resultado da simulação e dos dados observacionais, com pouca ou nenhuma argumentação explícita sobre

> o que, se alguma coisa, essas comparações indicam sobre a capacidade do modelo de fornecer evidências para hipóteses científicas específicas de interesse (2008b, p. 381).

Baseando-se explicitamente no trabalho de Mayo (1996), Parker argumenta que a epistemologia da simulação deveria, em vez disso, estar oferecendo um relato dos "erros canônicos" que podem surgir, bem como estratégias para investigar sua presença.

4.3 Verificação e validação

Praticantes de simulações, particularmente em contextos de engenharia, testes de armas e ciência climática, tendem a conceituar a ESC em termos de **verificação** e **validação**. Considera-se **verificação**, o processo de julgar se a saída da simulação se aproxima das soluções verdadeiras para as equações diferenciais do modelo original. Já, a **validação** é o processo que determinará se o modelo escolhido é uma representação suficientemente boa do sistema do mundo real para o objetivo da simulação. A literatura sobre verificação e validação de engenheiros e cientistas é enorme e está começando a receber alguma atenção dos filósofos.

A verificação pode ser dividida sob dois aspectos, a verificação da solução e verificação do código. O primeiro verifica se o resultado obtido a partir do algoritmo pretendido se aproxima das soluções verdadeiras para as equações diferenciais do modelo original. O último verifica se o código, conforme escrito executa o algoritmo pretendido. A verificação de código tem sido ignorada principalmente por filósofos da ciência; provavelmente porque foi visto mais como um problema da ciência da computação do que da ciência empírica - talvez isso seja um erro. Parte da verificação da solução consiste em comparar o resultado da computação com as soluções analíticas (as chamadas "soluções de Benchmark"). Embora esse método possa, é claro, ajudar a justificar os resultados de uma simulação computacional, é por si só inadequado, pois as simulações são frequentemente utilizadas, justamente, porque a solução analítica não está disponível para a solução que é de interesse. Outras técnicas indiretas estão disponíveis: a mais importante delas é provavelmente verificar se a taxa de saída computada converge para uma solução estável, à medida que o tempo e a resolução espacial da grade de discretização ficam mais refinados.

A principal estratégia de validação envolve a comparação dos resultados das computações do modelo com dados observáveis. Novamente, é claro, essa estratégia é limitada na maioria dos casos, onde simulações estão sendo executadas porque os dados observáveis são escassos. Mas estratégias complexas podem ser empregadas, incluindo a comparação da produção de subsistemas de uma simulação com experimentos relevantes (PARKER, 2013; OBERKAMPF; ROY, 2010).

Os conceitos de verificação e validação atraíram algumas críticas dos filósofos. Oreskes *et al*. (1994), em um artigo muito citado, criticou principalmente a terminologia, argumentando que "validade", em particular, é uma propriedade que se aplica apenas a argumentos lógicos e que, portanto, quando o termo é aplicado a modelos, pode levar ao excesso de confiança.

Winsberg (2010, 2018, p. 155) argumentou que a divisão conceitual entre verificação e validação pode ser enganosa, se for adotada para sugerir que existe um conjunto de métodos que podem, por si só, mostrar que resolvemos as equações adequadamente e que existe outro conjunto de métodos que podem, por si só, mostrar que temos as equações corretas. Ele também argumentou que é enganoso pensar que a epistemologia da simulação é claramente dividida em uma parte empírica (verificação) e uma parte matemática (referente à ciência da computação para a validação). Mas essa idéia enganosa geralmente segue a discussão sobre verificação e validação. Encontramos isso tanto no trabalho de práticos quanto de filósofos.

Tecnicamente Roy diz que: “A verificação lida com a matemática e aborda a correção da solução numérica para um determinado modelo. A validação, por outro lado, lida com a física e aborda a adequação do modelo na reprodução de dados experimentais. A verificação pode ser pensada como a solução correta das equações escolhidas, enquanto a validação está escolhendo as equações corretas em primeiro lugar” (ROY, 2005).

Alguns filósofos baseiam-se na distinção entre validação e verificação para apontarem a novidade filosófica da simulação. Levantamos inicialmente essa questão na seção 4.1, onde Frigg e Reiss argumentaram que a simulação não poderia ter características epistemologicamente novas, uma vez que continha dois componentes distintos: um que é idêntico à epistemologia da modelagem comum e outro que é inteiramente matemático. “Devemos distinguir duas noções diferentes de confiabilidade aqui, respondendo duas perguntas diferentes. Primeiro, as soluções que o computador fornece são próximas o suficiente das soluções reais (mas não disponíveis) e podem ser úteis? [...] Essa é uma questão puramente matemática e se enquadra na classe de problemas que acabamos de mencionar. Portanto, não há nada de novo aqui,

do ponto de vista filosófico, e a questão é de fato uma questão de processamento de números. Segundo, os modelos computacionais que são a base das simulações representam o sistema de destino corretamente? Ou seja, os resultados da simulação são válidos externamente? Essa é uma pergunta séria, mas é independente do primeiro problema e também se coloca em conexão com modelos que não envolvem matemática intratável e experimentos comuns" (FRIGG; REISS, 2009).

Mas a verificação e a validação não são, estritamente falando, tão claramente separáveis. Isso ocorre porque a maioria dos métodos de validação, por si só, são fracos demais para estabelecer a validade de uma simulação. E a maioria das equações presentes nos modelos e escolhidas para compor a simulação não são, em nenhum sentido direto, "as equações certas"; portanto estas não são as equações modelo que escolheríamos em um mundo ideal. Temos boas razões para pensar, em outras palavras, que existem equações modelo que seriam ferramentas empíricas melhores. As equações que escolhemos geralmente refletem um compromisso entre o que pensamos que melhor descreve os fenômenos e entre a tratabilidade computacional. Portanto, as equações escolhidas raramente são bem "validadas" por si próprias. Se quisermos entender por que os resultados da simulação são considerados credíveis, precisamos considerar a epistemologia da simulação como um todo integrado, não tão claramente dividido em verificação e validação - cada uma das quais, por si só, pareceria inadequada para a tarefa.

Ressaltamos que a verificação e validação não são atividades independentes e bem-sucedidas. Mas, por outro lado, não existem duas entidades independentes para as quais essas atividades possam ser direcionadas: um modelo escolhido para discretizar e um método para discretizá-lo. Uma vez que se reconhece que as equações a serem "resolvidas", às vezes, são escolhidas para cancelar erros de discretização, etc., essa distinção posterior é dificilmente mantida. A esse respeito, Lenhard (2007) propõe um exemplo muito bom envolvendo o operador de Arakawa. Então, o sucesso é obtido em simulação com um tipo de ajuste alternativo entre tentativa e erro, entre o modelo e o método de cálculo. E quando esse for o caso, é difícil até saber o que significa dizer que uma simulação é verificada e validada separadamente.

Não foi argumentado que a distinção entre validação e verificação não é útil, mas sim que os cientistas não devem exagerar pragmaticamente, em um ditado metodologicamente útil que deturpa os questionamentos necessários de sua própria prática. Colateralmente, o argumento de Frigg e Reiss pela ausência de novidade epistemológica na simulação falha por esse motivo. Não é "uma questão puramente matemática" se as soluções fornecidas pelo computador se aproximam o suficiente

das soluções reais (mas não disponíveis) para serem úteis. Pelo menos não a esse respeito: não é uma pergunta que possa ser respondida, como uma questão pragmática, usando inteiramente métodos matemáticos. E, portanto, é uma questão empírica/epistemológica que não surge na modelagem comum.

4.4 ESC e direito epistêmico

Uma das principais vertentes da epistemologia comum (não relacionada com a filosofia da ciência) enfatiza o modo condicional estabelecido para a possibilidade de conhecimento, segundo o qual confiamos em nossos sentidos e no testemunho de outras pessoas de uma maneira que não podemos justificar. Segundo Tyler Burge (1993, 1998), a crença nos resultados desses dois processos é garantida, mas não justificada. Pelo contrário, de acordo com Burge, temos **direito** a essas crenças. "[Temos] o direito de confiar, dentre outras coisas, na percepção, na memória, no raciocínio dedutivo e indutivo, e na [...] palavra dos outros" (1993, p. 458). As crenças que temos direito são aquelas que não são suportadas por evidências disponíveis para o sujeito que acredita, mas que tal sujeito é, no entanto, levado a crer.

Alguns trabalhos de ESC desenvolveram analogias entre a simulação computacional e os tipos de práticas de produção de conhecimento que Burge associa ao **direito** (*vide* BARBEROUSSE; VORMS, 2014; BEISBART, 2017). Isso é, de certa forma, uma conseqüência natural dos argumentos de Burge segundo os quais entendemos as provas computacionais dessa maneira (1998). As simulações computacionais são extremamente complexas, e geralmente é o resultado do trabalho epistêmico de um conjunto diversificado de cientistas e outros especialistas, e talvez o mais importante, epistemicamente opaco (HUMPHREYS, 2004). Por causa dessas características, Beisbart argumenta que é razoável tratar simulações computacionais da mesma maneira que tratamos nossos sentidos e o testemunho de outras pessoas: simplesmente como coisas que podem ser confiáveis, pois baseia-se na suposição de que tudo está funcionando satisfatoriamente (BEISBART, 2017).

Symons e Alvarado (2019) argumentam que há um problema fundamental com essa abordagem de ESC que está relacionado com um recurso de prova auxiliada por computador que foi crucial para o argumento original de Burge: a de ser um "transportador transparente". "É muito importante observar, por exemplo, que o relato de Burge sobre preservação de conteúdo e transporte transparente

exige que o destinatário já tenha motivos para não duvidar da fonte" (p. 13). Nesse sentido Symons e Alvarado apontam para muitas das propriedades das simulações computacionais, propostas por Winsberg (2010) e Ruphy (2015), que os modelos não possuem. Lenhard e Küster (2019) também são relevantes aqui, pois argumentam que existem muitos recursos da simulação computacional que os tornam difíceis de reproduzir e, portanto, comprometem parte da estabilidade necessária para que eles sejam transportadores transparentes. Por esses motivos e outros que estão relacionados com muitos dos recursos discutidos em 4.2 e 4.3, Symons e Alvarado argumentam que é implausível que devemos considerar a simulação computacional como uma prática epistêmica básica equivalente à percepção sensorial, memória, testemunho ou algo semelhante.

4.5 Abordagens pragmáticas da ESC

Outra abordagem da ESC é a tentativa de fundamentá-la em aspectos práticos do trabalho de modelagem e simulação. De acordo com essa visão, em outras palavras, o melhor relato que podemos dar dos motivos que nos levam a acreditar nos resultados dos estudos de simulação computacional é confiar nas habilidades práticas e nas habilidades dos modeladores que os utilizam. Um bom exemplo desse tipo de abordagem é o trabalho de Hubig e Kaminski (2017). O objetivo epistemológico de tal tipo de trabalho é identificar o local de nossa confiança nas simulações em aspectos práticos da modelagem e simulação, em vez de em quaisquer características dos próprios modelos. Resch et al (2017) argumentam que boa parte da razão pela qual devemos confiar nas simulações não é por causa das próprias simulações, mas sim por conta da interpretação dos resultados. Symons e Alvarado (2019) também criticam essa abordagem, argumentando que “parte da tarefa da epistemologia da simulação computacional é explicar a diferença entre a posição do cientista contemporâneo em relação às simulações epistemicamente opacas do computador [...]” (p. 7) e aqueles que acreditam na previsão mecânica. Considerações pragmáticas e epistêmicas, de acordo com Symons e Alvarado, coexistem e elas não são concorrentes possíveis para a explicação correta de nossa confiança nas simulações - as razões epistêmicas são as últimas que explicam e fundamentam as pragmáticas.

5. Simulação e Experiência

Os cientistas práticos às vezes descrevem estudos de simulação em termos experimentais. A conexão entre a simulação e o experimento provavelmente remontam a von Neumann, que ao defender pioneiramente o uso de computadores na física, observou que muitas experiências complexas tinham que ser conduzidas apenas para determinar fatos que, em princípio, devem ser derivados da teoria. Uma vez que a visão de von Neumann se tornou efetiva, e alguns desses experimentos começaram a ser substituídos por simulações, tornou-se algo natural considerá-los versões da experiência. Uma passagem representativa pode ser encontrada em um livro popular sobre simulação:

> Uma simulação que imita com precisão um fenômeno complexo contém muitas informações sobre o que é imitado. Variáveis como temperatura, pressão, umidade e velocidade do vento são avaliadas em milhares de pontos pelo supercomputador, quando simula o desenvolvimento de uma tempestade, por exemplo. Esses dados, que excedem em muito o que poderia ser obtido com o lançamento de uma frota de balões meteorológicos, revelam detalhes íntimos do que está acontecendo na nuvem de tempestade (KAUFMANN; SMARR, 1993, p. 4).

A idéia de experimentos "*in* silício" se torna ainda mais plausível quando uma simulação é projetada para representar o que ocorre em um sistema com várias intervenções possíveis: O que aconteceria com o clima global se **X** quantidade de carbono fosse adicionada à atmosfera? O que acontecerá com esta asa de avião se estiver sujeita a tal e qual tensão? Como os padrões de tráfego mudariam se um acesso fosse adicionado nesse local?

Os filósofos, consequentemente, começaram a considerar em que medida, se for o caso, simulações computacionais são como experimentos e em que sentido não são. Uma questão relacionada diz respeito a determinar quando um processo que envolve fundamentalmente simulação computacional pode ser considerado um experimento (PARKER, 2017). Algumas visões emergiram na literatura especializada, defendendo e criticando duas teses distintas:

A **tese da identidade**. Os estudos a partir de simulação computacional são literalmente instâncias de experimentos.
A **tese da dependência epistemológica**. A tese da identidade seria (se fosse verdade) uma boa razão (versão fraca), ou a melhor razão (versão forte), ou a única razão (versão mais forte; é uma condição necessária) para acreditar que as simulações podem fornecer garantias para crença nas hipóteses que sustentam determinado estudo. Uma conseqüência da versão mais forte é que se a tese da identidade for verdadeira, há razões para acreditar que as simulações podem garantir a credibilidade de hipóteses.

A idéia central por trás da tese da dependência epistemológica é que os experimentos são as entidades canônicas que desempenham um papel relevante na garantia de nossa crença nas hipóteses científicas, e que, portanto, o grau em que devemos pensar que as simulações também podem desempenhar um papel na garantia de tais crenças depende da extensão a partir da qual podem ser identificados como um tipo de experimento.

Pode-se encontrar filósofos defendendo a tese da identidade desde Humphreys (1995) e Hughes (1999). E há pelo menos apoio implícito à versão (mais forte) da tese da dependência epistemológica em Hughes. O argumento explícito mais antigo a favor da tese da dependência epistemológica, no entanto, está em Norton e Suppe (2001). Segundo tais autores, as simulações podem justificar a crença em hipóteses científicas precisamente porque são literalmente experimentos. Estes justificam-se por si mesmos em que sentido são experimentos e como tudo deve funcionar. De acordo com Norton e Suppe, uma simulação válida é aquela em que certas relações formais (o que elas chamam de 'realização') se mantêm entre um modelo base, o próprio sistema físico modelado e o computador executando o algoritmo. Quando as condições adequadas são atendidas, “uma simulação pode ser usada como um instrumento para sondar ou detectar fenômenos do mundo real. Dados empíricos sobre fenômenos reais são produzidos sob condições de controle experimental” (p. 73).

Um problema com tal abordagem é que as condições formais são muito rígidas. É improvável que existam muitos exemplos reais de simulações computacionais que atendam aos rígidos padrões estabelecidos. A simulação é quase sempre idealizada e aproximativa. Portanto, se simulações são experimentos, provavelmente

não é da maneira que Norton e Suppe imaginavam.

De maneira mais geral, a tese da identidade interessou estudiosos de outros ramos de pesquisa.

Gilbert e Troitzsch argumentaram que "[a] principal diferença é que, enquanto em um experimento, alguém está controlando o objeto de interesse real (por exemplo, em um experimento de química, os produtos químicos sob investigação), em uma simulação está experimentando um modelo em vez do fenômeno em si." (GILBERT; TROITZSCH, 1999, p. 13). Mas isso não parece certo. Muitos pesquisadores (GUALA, 2002, 2008; MORGAN, 2003; PARKER, 2009a; WINSBERG, 2009a) apontaram problemas com tal afirmação. Se Gilbert e Troitzsch querem dizer que os simulacionistas manipulam modelos no sentido de objetos abstratos, então a afirmação é difícil de entender, como manipulamos uma entidade abstrata? Se, por outro lado, eles simplesmente pretendem apontar para o fato de que o objeto físico que os simulacionistas manipulam, um computador digital, não é o objeto de interesse real, então, não está claro por que isso difere dos experimentos.

Não é necessariamente verdadeiro, afirmar que experimentos reais sempre manipulem exatamente os objetos de interesse. De fato, tanto em experimentos reais quanto em simulações, existe uma relação complexa entre o que é manipulado na investigação, por um lado, e os sistemas do mundo real, que são os objetos da investigação, por outro. Portanto, em casos de experimento e simulação, é necessário um argumento que sustente a "validade externa" da investigação, para estabelecer se o que é aprendido sobre o sistema que está sendo manipulado é aplicável ao sistema de interesse. Mendel, por exemplo, manipulou plantas de ervilha, mas estava interessado em aprender sobre o fenômeno da hereditariedade em geral. A idéia de um **organismo modelo** em biologia torna isso uma idéia perspícua. Realizamos experimentos com *Caenorhabditis elegans* porque estamos interessados em entender como o organismo em geral utiliza genes para controlar o desenvolvimento e a genealogia. Colocamos *Drosophila melanogaster* sob experimentação, porque fornece um modelo útil de mutações e herança genética. Mas a ideia não se limita à biologia. Galileu experimentou planos inclinados porque estava interessado em saber como os objetos caem e como eles se comportariam na ausência de forças interferentes - fenômenos que os experimentos em planos inclinados nem sequer instanciaram.

Claramente, essa visão sobre experimentos é contestável. É fato que, frequentemente, os experimentalistas inferem algo sobre um sistema distinto do sistema em que interferem. No entanto, não está claro se essa inferência é parte apropriada do experimento original. Peschard (2010) faz uma crítica nesse sentido

e, portanto, pode ser visto como um defensor de Gilbert e Troitzsch. Peschard argumenta que a suposição fundamental de seus críticos, que na experimentação, assim como na simulação, o que é manipulado é um sistema que substitui um sistema de destino, está confuso. Peschard argumenta que confunde o **alvo epistêmico** de um experimento com sua **motivação epistêmica**. A autora discorre que, embora a motivação epistêmica para realizar experimentos com *C. elegans* possa ser bastante abrangente, o alvo epistêmico adequado para qualquer experimento desse tipo é o próprio verme. Em uma simulação, de acordo com Peschard, no entanto, o alvo epistêmico nunca é o próprio computador digital. Assim, é distinto do experimento, segundo ela, na medida em que seu objetivo epistêmico (em oposição a apenas sua motivação epistêmica) é distinto do objeto que está sendo manipulado. Roush (2017) também pode ser visto como um defensor da linhagem Gilbert e Troitzsch, mas apela à mesmice de tipos naturais como o recurso crucial que separa experimentos e simulações. Outros oponentes da tese de identidade incluem Giere (2009) e Beisbart e Norton (2012).

Não está claro analisar essa disputa que parece se fundamentar principalmente em torno de uma diferença de ênfase. Pode-se enfatizar a diferença entre experimento e simulação, seguindo Gilbert e Troitzsch e Peschard, insistindo que os experimentos nos ensinam primeiro sobre seus alvos epistêmicos e apenas secundariamente permitem inferências sobre o comportamento de outros sistemas (ou seja, experimentos utilizando vermes nos ensinam, em primeira instância, sobre vermes, e apenas secundariamente nos permitem fazer inferências sobre o controle genético de maneira mais geral).

Isso os tornaria conceitualmente diferentes das simulações em computador, que não se pensa em nos ensinar, na primeira instância, sobre o comportamento dos computadores, e somente na segunda instância sobre tempestades, galáxias ou qualquer outra coisa.

Ou pode-se enfatizar a **similaridade** da maneira oposta. Pode-se enfatizar o grau em que os alvos experimentais são sempre escolhidos como substitutos do que é realmente interessante. Morrison (2009) é provavelmente a defensora mais forte ao enfatizar esse aspecto da semelhança entre experimento e simulação. Ela argumenta que a maioria das práticas experimentais, e de fato a maioria das práticas de medição, envolvem os mesmos tipos de práticas de modelagem que as simulações. Segundo Peschard, isso não é nada além de um debate sobre nomenclatura - e talvez um apelo ao uso comum da linguagem pelos cientistas; nem sempre o tipo mais convincente de argumento - nos impediria de dizer que o alvo epistêmico de

uma simulação de tempestade é o computador e que a tempestade é apenas a motivação epistêmica para estudar o computador.

Seja como for, muitos filósofos interessados no estudo das simulações, incluindo os discutidos nesta seção, escolheram o último caminho, em parte como uma maneira de chamar a atenção para as maneiras pelas quais a mensagem oculta por trás da alegação citada de Gilbert e Troitzsch oferece uma imagem excessivamente simplista do ato da experiência. Parece excessivamente simplista considerar que o experimento atinge **diretamente** o mundo, enquanto a simulação funcionaria de modo oposto. E esta é a abordagem que parece obtermos da citação de Gilber e Troitzsch. Já a concepção elaborada por Peschard envolve uma distinção entre objetivos e motivações epistêmicas, e auxilia a minimizar preocupações sem nos levar a pensar que simulação e experimento são **exatamente iguais**, nesse sentido.

Ainda assim, apesar de rejeitar a caracterização de Gilbert e Troitzsch que diferencia simulação de experimento, Guala e Morgan rejeitam a tese da identidade. Com base no trabalho de Simon (1969), Guala argumenta que as simulações diferem fundamentalmente das experiências, pois o objeto de manipulação em uma experiência apresenta uma semelhança material com o objeto de interesse, mas em uma simulação, a semelhança entre objeto e o objetivo simulacional é meramente formal. Curiosamente, embora Morgan aceite esse argumento contra a **tese da identidade**, a autora parece se apegar a uma versão da **tese da dependência epistemológica**. Morgan argumenta, em outras palavras, que a diferença entre experimentos e simulações identificadas por Guala implica que as simulações são epistemologicamente inferiores aos experimentos reais - ou seja, as simulações possuem intrinsecamente menos poder para garantir a crença em hipóteses sobre o mundo real, pois não são experimentos.

Uma defesa do poder epistêmico das simulações contra o argumento de Morgan (2002) poderia ser sustentada ao considerarmos a relevância da tese da identidade ou ao rejeitarmos a tese da dependência epistemológica. De um lado, parece haver dois problemas com o argumento de Guala (2002) contra a tese da identidade. A primeira é que a noção de similaridade material é muito fraca, e a segunda é que a noção de mera similaridade formal é muito vaga para defender o poder epistêmico das simulações. Considere, por exemplo, o fato de que não é incomum, nas ciências da engenharia, usar métodos de simulação para estudar o comportamento de sistemas fabricados a partir de silício. O engenheiro deseja aprender sobre as propriedades de diferentes possibilidades de design para um dispositivo de silício, e desenvolve um modelo computacional do dispositivo a fim

de executar uma simulação de seu comportamento em um computador digital. Existem semelhanças materiais entre o objeto e a instanciação computacional da simulação, além do mais algumas das mesmas causas materiais estão em ação, no processador central do computador e no dispositivo de silício em estudo. Na abordagem de Guala, tal semelhança figura como um exemplo de um experimento real, o que parece ser errado. As peculiaridades deste exemplo ilustram bastante o problema, mas este é de fato bastante geral: quaisquer dois sistemas têm algumas semelhanças materiais entre si e algumas diferenças.

Parke (2014) argumenta contra a tese da dependência epistemológica, questionando duas premissas, segundo a autora, são basilares: os experimentos geram maior poder inferencial que as simulações, e que as simulações não podem nos surpreender da mesma maneira que os experimentos.

Por outro lado, a ideia de que a existência de uma similaridade formal entre duas entidades materiais é conceitualmente confusa. Dadas duas entidades suficientemente complexas, existem muitas maneiras pelas quais elas são formalmente idênticas, e até mesmo semelhantes. Existem também maneiras pelas quais são formal e completamente diferentes. Agora, podemos falar livremente e dizer que duas coisas têm uma semelhança formal, mas o que realmente queremos dizer é que nossas melhores representações formais das duas entidades têm semelhanças. De qualquer forma, parece haver bons motivos para rejeitar os motivos de Gilbert e Troitzsch e Morgan e Guala para distinguir experimentos e simulações.

Voltando à defesa do poder epistêmico das simulações, há também motivos para rejeitar a tese da dependência epistemológica. Como Parker (2009a) aponta, tanto no experimento quanto na simulação, podemos constatar semelhanças relevantes entre simulações computacionais e sistemas de destino, e é isso que importa. Quando o conhecimento relevante está em vigor, uma simulação pode fornecer um conhecimento mais confiável de um sistema do que um experimento. Uma simulação computacional do sistema solar, com base em modelos sofisticados da dinâmica celeste, produzirá melhores representações das órbitas dos planetas do que qualquer experimento.

Parke (2014) argumenta contra a tese da dependência epistemológica, minando duas premissas que ela acredita que a sustentam: primeiro, que os experimentos geram maior poder inferencial que as simulações, e segundo, que as simulações não podem nos surpreender da mesma maneira que os experimentos. O argumento de que simulações não podem nos surpreender é defendido por Morgan (2005). Contra Morgan, Parke argumenta que os simulacionistas são

freqüentemente surpreendidos por suas simulações, tanto porque não são oniscientes em termos computacionais, como porque nem sempre são os únicos criadores dos modelos e códigos que usam. Ela argumenta, além disso, que “diferenças nos estados epistêmicos do pesquisador, por si só, parecem ser os motivos errados para rastrear uma distinção entre experimento e simulação” (p. 258). Adrian Curry (2017) defende a intuição original de Morgan, com duas alterações amigáveis. Curry argumenta que a distinção que Morgan realmente buscava era entre dois tipos diferentes de surpresa e, em particular, qual é a fonte da surpresa: surpresa por colocar conhecimento teórico em contato com o mundo é uma característica do experimento. Ele também define com mais cuidado a surpresa de maneira não psicológica, de modo que seja uma "qualidade cuja realização constitui genuíno progresso epistêmico" (p. 640).

6. Simulação computacional e a estrutura das teorias científicas

Paul Humphreys (2004) argumentou que as simulações computacionais têm implicações profundas na nossa compreensão da estrutura das teorias; pois revelam inadequações com as visões semântica e sintática das teorias científicas. Esta alegação provocou um forte ataque de Frigg e Reiss (2009). Estes argumentam que o fato de um modelo admitir ou não uma solução analítica não tem influência sobre como ele se relaciona com o mundo. Para ilustrar tal abordagem, os autores utilizam o pêndulo duplo. Sendo o ponto de apoio interno do pêndulo fixo ou não (um fato que determinará se o modelo relevante é analiticamente solucionável) não tem influência na semântica dos elementos do modelo. A partir disso, os autores alegam que a semântica de um modelo, ou como ele se relaciona com o mundo, não é afetada se tal modelo é ou não analiticamente solucionável.

A abordagem, entretanto, não é a mesma quando analisamos o que Humphreys aponta. Afinal, as visões sintática e semântica das teorias não eram apenas relatos de como nossas representações científicas abstratas se relacionam com o mundo. Mais particularmente, não eram histórias sobre a relação entre modelos particulares e o mundo, mas sim sobre a relação entre teorias e o mundo, e o papel, se tem algum, que os modelos desempenhavam nessa relação.

As abordagens anteriores revelam o aspecto filosoficamente interessante quando se trata de teorização científica. A visão sintática sugeria que a prática científica poderia ser adequadamente reconstruída de forma racional se as teorias

fossem consideradas sistemas axiomáticos e, mais importante, que a dedução lógica fosse um ideal regulador útil para pensar sobre como as inferências da teoria para o mundo são feitas. A visão sintática, também tornou bastante nítida a hipótese de que a modelagem desempenha apenas um papel **heurístico** na ciência (essa era uma característica da visão sintática das teorias contra a qual Frederick Suppe, um de seus críticos mais fervorosos, costumava combater). As teorias em si não se relacionavam com modelos, e as teorias podiam ser comparadas **diretamente** ao mundo, sem qualquer papel importante para modelar.

A visão semântica das teorias, por outro lado, enfatizava um papel importante para os modelos, mas também defendia que as teorias eram entidades não linguísticas. Exortava os filósofos a não se distrair com as contingências da forma particular de expressão linguística em que uma teoria pode ser encontrada, digamos, em um determinado livro didático.

Simulações computacionais, entretanto, parecem ilustrar que ambas as teorias possuem falhas. Era equivocado considerar que a dedução lógica figurava enquanto ferramenta adequada para reconstruir racionalmente o processo de aplicação da teoria. Simulações computacionais mostram que existem métodos de aplicação da teoria que superam amplamente o poder inferencial da dedução lógica. O espaço de soluções, por exemplo, que está disponível por dedução lógica da teoria dos fluidos é microscópico se comparado ao espaço de aplicações que pode ser explorado por simulação computacional.

Por outro lado, as simulações computacionais parecem revelar que, como Humphreys (2004) insistiu, a sintaxe é importante. No entanto, é equivocado sugerir, como fazia a perspectiva semântica, que a forma linguística particular em que uma teoria científica é expressa, é filosoficamente desinteressante. A sintaxe da expressão da teoria terá um efeito profundo sobre quais inferências podem ser extraídas dela, quais tipos de idealizações funcionarão bem, etc. Humphreys colocou a questão da seguinte forma: "a representação sintática específica usada é muitas vezes crucial para a solvabilidade das equações da teoria" (HUMPHREYS, 2009, p. 620). A teoria dos fluidos pode ser usada para enfatizar este ponto: a escolha da forma de expressão dessa teoria, em euleriano ou na forma lagrangiana, afetará profundamente o que, na prática, podemos calcular e como; afetarão também quais idealizações, aproximações e técnicas de cálculo serão eficazes e confiáveis em determinadas circunstâncias. Portanto, a epistemologia da simulação computacional precisa ser sensível à formulação sintática particular de uma teoria e à quão bem essa formulação particular foi credenciada. Portanto, parece correto enfatizar, como

fez Humphreys (2004), que as simulações computacionais revelaram inadequações tanto com a teoria sintática quanto semântica.

7. Emergência

Paul Humphreys (2004) e Mark Bedau (1997, 2011) argumentaram que os filósofos interessados no tópico da emergência podem aprender muito analisando simulações computacionais.

A conexão entre emergência e simulação foi talvez melhor articulada por Bedau em seu trabalho de 2011. Nele, Bedau argumentou que qualquer concepção de emergência deve atender às características relevantes que podem explicar como o todo depende de suas partes e como o todo é independente de suas partes. Nesse sentido, o autor aponta que os filósofos muitas vezes focam no que ele chama de emergência "forte", que diz respeito à causação descendente que é irredutível em princípio. No entanto, Bedau defende que essa concepção é equivocada. Em vez disso, ele se concentra no que chama de emergência "fraca", que **permite** a redutibilidade de todos a partes em **princípio**, mas não na **prática**. Os sistemas que produzem propriedades emergentes são meros mecanismos, mas estes são muito complexos (eles têm muitas partes interagindo independentemente). Como resultado, não há como descobrir exatamente o que acontecerá dado um conjunto específico de condições iniciais e de limites, exceto para "rastrear a teia causal". É aqui que surge a conexão com a simulação computacional. Propriedades fracamente emergentes são características de sistemas complexos na natureza. Essa situação é também característica de simulações computacionais complexas nas quais não há como prever o que farão, sendo apenas possível deixá-las funcionar. A emergência fraca explica, de acordo com Bedau, por que as simulações computacionais desempenham um papel central na ciência dos sistemas complexos. A melhor maneira de compreender e prever como sistemas complexos reais se comportam é simulando-os, de modo a rastrear a teia micro-causal e analisar o que ocorre.

8. Ficções

É claro que os modelos envolvem idealizações. Mas tem sido argumentado que alguns tipos de idealização, que desempenham um papel especialmente

proeminente nos tipos de modelagem envolvidos na simulação computacional, são especiais - a ponto de merecerem o título de "ficção". Nesta seção serão discutidas as tentativas de definir ficções e explorar seu papel na simulação computacional.

Existem duas linhas teóricas distintas acerca do papel das ficções na ciência. Segundo uma delas, todos os modelos são ficções. Essa linha de pensamento é motivada por considerar o papel, por exemplo, do “pêndulo ideal” na ciência. Os cientistas costumam fazer afirmações sobre esses tipos de entidades (por exemplo, “o pêndulo ideal tem um período proporcional à raiz quadrada de seu comprimento), mas elas não são encontradas em nenhum lugar do mundo real; portanto, devem ser entidades fictícias. Esta linha de argumento sobre entidades ficcionais na ciência não se conecta de nenhuma maneira especial com simulações computacionais - os leitores interessados neste tópico devem consultar a entrada sobre representação científica [no prelo].

Por outro lado, há a linha teórica acerca das ficções, que está preocupada com a questão que dizem respeito sobre quais tipos de representações na ciência devem ser consideradas ficcionais. Aqui, a preocupação não é tanto sobre a ontologia de entidades do modelo científico, mas sobre o caráter representacional das várias entidades postuladas. Nesse sentido, Winsberg (2009c) argumentou que as ficções têm uma conexão especial com as simulações computacionais. Ou melhor, que algumas simulações computacionais contêm elementos que tipificam satisfatoriamente o que podemos chamar de representações ficcionais na ciência, mesmo que essas representações não estejam exclusivamente presentes nas simulações.

Winsberg aponta que a primeira concepção de ficção, mencionada acima, que torna "qualquer representação que contradiz a realidade uma ficção" (p. 179), não corresponde ao nosso uso comum do termo: um mapa mal feito não é ficção. O autor então propõe uma definição alternativa: a não-ficção é oferecida como um guia “bom o suficiente” para alguma parte do mundo (p. 181); ficção, não. Mas a definição precisa ser refinada. Veja a fábula do gafanhoto e da formiga. Embora a fábula ofereça lições sobre como o mundo é, ainda é ficção porque é "um guia útil para a forma como o mundo é em algum sentido geral", em vez de um guia específico para a forma como uma parte do mundo é, isto é, o “alvo representacional prima facie”, é um gafanhoto cantor e uma formiga trabalhadora. As não-ficções, por outro lado, “apontam para uma determinada parte do mundo” e é um guia para aquela parte do mundo (p. 181).

Esses tipos de componentes fictícios de modelos são exemplificados paradigmaticamente em certas simulações computacionais. Dois exemplos são o

"átomo de silogênio" e a "viscosidade artificial". Os átomos de silogênio aparecem em certos modelos nano mecânicos em fendas de silício, uma espécie do tipo de modelos multiescala que combinam a mecânica quântica e a molecular mencionada na seção 2.3. Os modelos de propagação de fenda de silício contendo silogênio descrevem a própria fenda usando a mecânica quântica e a região ao redor utilizando a dinâmica molecular clássica. Para abarcar as estruturas de modelagem nas duas regiões distintas, o limite é tratado como se contivesse átomos de "silogênio", que possui uma mistura das propriedades do silício e do hidrogênio. Os átomos de silogênio são ficções. Eles não são uma descrição "boa o suficiente" dos átomos do limite, seus alvos representacionais *prima facie*. Mas eles são utilizados para que se possa esperar que o modelo geral seja adequado. Assim, o modelo geral não é uma ficção, mas um de seus componentes é. A viscosidade artificial é um exemplo semelhante. Fluidos que sofrem choques abruptos são difíceis de modelar em uma grade computacional porque tal choque ocorre dentro de uma única célula da grade e não pode ser resolvido por algoritmo. A viscosidade artificial é uma técnica que simula o fluido como altamente viscoso, uma ficção, em que o choque ocorre, de modo que este seja menos abrupto e ecoe em várias células. Obter a viscosidade e, portanto, a intensidade do choque, auxilia o modelo geral funcionar "bem o suficiente". Novamente, o modelo geral do fluido não é uma ficção, mas sim um guia confiável o suficiente para o comportamento do fluido. Mas o componente chamado viscosidade artificial é uma ficção, não está sendo usado para modelar o choque de maneira confiável. Ele está sendo incorporado a uma estrutura de modelagem maior, de modo a torná-la "confiável o suficiente".

A situação acima atraiu dois tipos de críticas. Toon (2010) argumentou que essa definição de ficção é muito estreita. Ele exemplifica dramas históricos como *Eu, Claudius* e *A lista de Schindler*, como ficções, apesar do fato de que "elas são oferecidas como guias 'bons o suficiente' para as pessoas, acerca dos lugares e eventos em certos aspectos" (pp. 286-287). Toon, presumivelmente, apóia uma concepção mais ampla do papel das ficções na ciência, segundo a qual elas não desempenham um papel particularmente proeminente ou elevado na simulação computacional.

Gordon Purves (no prelo) argumenta que existem exemplos de ficções em modelos computacionais (seu exemplo são as chamadas "fendas imaginárias"), e em outros âmbitos, que não atendem aos requisitos estritos discutidos acima. Ao contrário de Toon, Purves intenciona delinear elementos de modelagem ficcionais dos não-ficcionais. Sua principal crítica é ao critério de ficcionalidade em termos de normas sociais de uso, e Purves argumenta que devemos ser capazes de decidir

se alguma peça de modelagem é ou não uma ficção na ausência de tais normas. Assim, deseja-se encontrar uma caracterização intrínseca de uma ficção científica. A proposta de Purves toma como constitutivas as ficções modelo que deixam de ter a característica que Laymon (1985) chamou de "improbabilidade parcial" (IP). IP é uma característica de muitos modelos idealizados; segundo a qual, conforme se desidealiza, o modelo se torna cada vez mais preciso. Mas, à medida que se desidealiza um átomo de silogênio, não se obtém uma simulação cada vez mais precisa de uma fenda de silício. Mas Purves considera essa falha da IP como constitutiva de uma ficção, ao invés de meramente sintomática dela.

Referências

BARBEROUSSE, A.; LUDWIG, P. **Models as Fictions in Fictions in Science. Philosophical Essays in Modeling and Idealizations**. London: Routledge, 2009, p. 56-73.

BARBEROUSSE, A.; VORMS, M. About the warrants of computer-based empirical knowledge. **Synthese**, v. 191, n.15, p. 3595-3620, 2014.

BEDAU, M. A. Weak emergence and computer simulation. *In:* HUMPHREYS, P.; IMBERT, C. (ed.). **Models, Simulations, and Representations**. New York: Routledge, 2011, p. 91-114.

BEDAU, M. A. Weak Emergence. **Noûs**, Sup. 11, n. 31, p. 375-399, 1997.

BEISBART, C.; NORTON, J. Why Monte Carlo Simulations are Inferences and not Experiments. *In:* **International Studies in Philosophy of Science**, v. 26, p. 403-422, 2012.

BEISBART, C. Advancing knowledge through computer simulations? A socratic exercise. *In:* RESCH, M.; KAMINSKI, A.; GEHRING, P. (ed.). **The Science and Art of Simulation** (Volume I). Cham: Springer, 2017, p. 153-174.

BURGE, T. Content preservation. **The Philosophical Review**, v. 102, n. 4, p. 457-488, 1993.

BURGE, T. Computer proof, apriori knowledge, and other minds: The sixth philosophical perspectives lecture. **Noûs**, v. 32, n. S12, p. 1-37, 1998.

CURRIE, A. The argument from surprise. **Canadian Journal of Philosophy**, v. 48, n. 5, p. 639-661, 2018.

DARDASHTI, R.; THEBAULT, K.; WINSBERG, E. Confirmation via analogue simulation: what dumb holes could tell us about gravity. *In:* **British Journal**

for the Philosophy of Science, v. 68, n. 1, p. 55-89, 2015.

DARDASHTI, R.; HARTMANN, S.; THEBAULT, K.; WINSBERG, E. Hawking radiation and analogue experiments: A Bayesian analysis. **Studies in History and Philosophy of Modern Physics**, v. 67, p. 1-11, 2019.

EPSTEIN, J.; AXTELL, R. **Growing artificial societies**: Social science from the bottom-up. Cambridge: MIT Press, 1996.

EPSTEIN, J. Agent-based computational models and generative social science. **Complexity**, v. 4, n. 5, p. 41- 57, 1999.

FRANKLIN, A. **The Neglect of Experiment.** Cambridge: Cambridge University Press, 1996.

FRANKLIN, A. **The Epistemology of Experiment**. The Uses of Experiment. Cambridge: Cambridge University Press, 1989, p. 437-460.

FRIGG, R.; REISS, J. The philosophy of simulation: Hot new issues or same old stew. **Synthese**, v. 169, p. 593-613, 2009.

GIERE, R. N. Is Computer Simulation Changing the Face of Experimentation? **Philosophical Studies**, v. 143, p. 59-62, 2009.

GILBERT, N.; TROITZSCH, K. **Simulation for the Social Scientist**. Philadelphia: Open University Press, 1999.

GRÜNE-YANOFF, T. Bounded Rationality. **Philosophy Compass**, v. 2, n. 3, p. 534-563, 2007.

GRÜNE-YANOFF, T.; WEIRICH, P. Philosophy of Simulation. Simulation and Gaming. **An Interdisciplinary Journal**, v. 41, n. 1, p. 1-31, 2010.

GUALA, F. **Models, Simulations, and Experiments**. **Model-Based Reasoning**: Science, Technology, Values. New York: Kluwer, 2002, p. 59-74.

GUALA, F. Paradigmatic Experiments: The Ultimatum Game from Testing to Measurement Device. **Philosophy of Science**, v. 75, p. 658-669, 2008.

HACKING, I. **Representing and Intervening**: Introductory Topics in the Philosophy of Natural Science. Cambridge: Cambridge University Press, 1983.

HACKING, I. On the Stability of the Laboratory Sciences. **The Journal of Philosophy**, v. 85, p. 507-515, 1988.

HACKING, I. Do Thought Experiments have a Life of Their Own? PSA (Volume 2). *In:* FINE, A.; FORBES, M.; OKRUHLIK, K. (ed.). East Lansing. **The Philosophy of Science Association**, p. 302-310, 1992.

HARTMANN, S. The World as a Process: Simulations in the Natural and Social Sciences. *In:* HEGSELMANN, R. et al. (ed.). **Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View**.

Dordrecht: Kluwer, 1996, p. 77-100.

HUBIG, C.; KAMINSKI, A. Outlines of a pragmatic theory of truth and error in computer simulation. *In:* RESCH, M.; KAMINSKI, A.; GEHRING, P. (eds.). **The Science and Art of Simulation.** v. 1. Cham: Springer, p. 121-136, 2017.

HUGHES, R. The Ising Model, Computer Simulation, and Universal Physics. *In:* MORGAN, M.; MORRISON, M. (ed.). **Models as Mediators.** Cambridge: Cambridge University Press, 1999.

HUGGINS, E. M.; SCHULTZ, E. A. San Francisco bay in a warehouse. **Journal of the Institute of Environmental Sciences and Technology**, v. 10, n. 5, p. 9-16, 1967.

HUMPHREYS, P. Computer Simulation. *In:* FINE, A.; FORBES, M.; WESSELS, L. (ed.). PSA 1990 v. 2. East Lansing. **The Philosophy of Science Association**, p. 497-506, 1990.

HUMPHREYS, P. Computational science and scientific method. **Minds and Machines**, v. 5, n. 1, p. 499-512, 1995.

HUMPHREYS, P. **Extending ourselves**: Computational science, empiricism, and scientific method. New York: Oxford University Press, 2004.

HUMPHREYS, P. The philosophical novelty of computer simulation methods. **Synthese**, v. 169, p. 615-626, 2009.

KAUFMANN, W. J.; SMARR, L. L. **Supercomputing and the Transformation of Science**. New York: Scientific American Library, 1993.

LAYMON, R. Idealizations and the testing of theories by experimentation. *In:* ACHINSTEIN, P.; HANNAWAY, O. (ed.). **Observation, Experiment and Hypothesis in Modern Physical Science**. Cambridge: MIT Press, 1985, p. 147-173.

LENHARD, J. Computer simulation: The cooperation between experimenting and modeling. **Philosophy of Science**, v. 74, p. 176-194, 2007.

LENHARD, J. Calculated Surprises: A Philosophy of Computer Simulation. Oxford: Oxford University Press, 2019.

LENHARD, J.; KÜSTER, U. **Minds & Machines**, v. 29, n. 19, 2019.

MORGAN, M. Experiments without material intervention: Model experiments, virtual experiments and virtually experiments. *In:* RADDER, H. (ed.). **The Philosophy of Scientific Experimentation.** Pittsburgh: University of Pittsburgh Press, 2003, p. 216-235.

MORRISON, M. Models, measurement and computer simulation: The changing face of experimentation. **Philosophical Studies**, v. 143, p. 33-57, 2012.

NORTON, S.; SUPPE, F. Why atmospheric modeling is good science. *In:* MILLER, C.; EDWARDS, P. (ed.). **Changing the Atmosphere**: Expert Knowledge and Environmental Governance. Cambridge, MA: MIT Press, 2001, p. 88-133.

OBERKAMPF, W.; ROY, C. **Verification and Validation in Scientific Computing.** Cambridge: Cambridge University Press, 2010.

ORESKES, N.; SHRADER-FRECHETTE, K.; BELITZ, K. Verification, Validation and Confirmation of Numerical Models in the Earth Sciences. **Science**, v. 263, n. 5147, p. 641-646, 1994.

PARKE, E. Experiments, Simulations, and Epistemic Privilege. **Philosophy of Science**, v. 81, n. 4, p. 516-536, 2014.

PARKER, W. Franklin, Holmes and the Epistemology of Computer Simulation. **International Studies in the Philosophy of Science**, v. 22, n. 2, p. 165-183, 2008a.

PARKER, W. Computer Simulation through an Error-Statistical Lens. **Synthese**, v. 163, n. 3, p. 371-384, 2008b.

PARKER, W. Does Matter Really Matter? Computer Simulations, Experiments and Materiality. **Synthese**, v. 169, n. 3, p. 483-496, 2009a.

PARKER, W. Computer Simulation. *In:* PSILLOS, S.; CURD, M. (ed.). **The Routledge Companion to Philosophy of Science.** 2. ed. London: Routledge, 2013.

PARKER, W. Computer Simulation, Measurement, and Data Assimilation. **British Journal for the Philosophy of Science**, v. 68, n. 1, p. 273-304, 2017.

PESCHARD, I. Modeling and Experimenting. *In:* HUMPHREYS, P.; IMBERT, C. (ed.). **Models, Simulations, and Representations.** London: Routledge, 2010, p. 42-61.

PRIMEIRO, G., 2019. "A Minimalist Epistemology for Agent-Based Simulations in the Artificial Sciences," Minds and Machines, 29(1): 127–148.

PURVES, G. M. Finding truth in fictions: identifying non-fictions in imaginary cracks. **Synthese,**

RESCH, M. M.; KAMINSKI, A.; GEHRING, P. (ed.). **The science and art of simulation I**: Exploringunderstanding-knowing. Berlin: Springer, 2017.

ROUSH, S. The epistemic superiority of experiment to simulation. **Synthese**, v. 169, p. 1-24, 2015.

ROY, S. Recent advances in numerical methods for fluid dynamics and heat transfer. **Journal of Fluid Engineering**, v. 127, n. 4, p. 629-630, 2005.

RUPHY, S. Computer simulations: A new mode of scientific inquiry? *In:* HANSEN, S. O. (ed.). **The Role of Technology in Science**: Philosophical Perspectives,

Dordrecht: Springer, 2015, p. 131-149.
SCHELLING, T. C. Dynamic Models of Segregation. **Journal of Mathematical Sociology**, v. 1, p. 143-186, 1971.
SIMON, H. **The Sciences of the Artificial**. Boston: MIT Press, 1969.
SYMONS, J.; ALVARADO, R. Epistemic Entitlements and the Practice of Computer Simulation. **Minds and Machines**, v. 29, n. 1, p. 37-60, 2019.
TOON, A. Novel Approaches to Models. **Metascience**, v. 19, n. 2, p. 285-288, 2010.
TRENHOLME R. Analog Simulation. **Philosophy of Science**, v. 61, p. 115-131, 1994.
UNRUH, W. G. Experimental black-hole evaporation? **Physical Review Letters**, v. 46, n. 21, p. 1351-1353, 1981.
WINSBERG, E. **Philosophy and Climate Science.** Cambridge: Cambridge University Press, 2018.
WINSBERG, E. **Science in the Age of Computer Simulation**. Chicago: The University of Chicago Press, 2010.
WINSBERG, E. A Tale of Two Methods. **Synthese**, v. 169, n. 3, p. 575-592, 2009a.
WINSBERG, E. Computer Simulation and the Philosophy of Science. **Philosophy Compass**, v. 4/5, p. 835-845, 2009b.
WINSBERG, E. A Function for Fictions: Expanding the scope of science. *In:* SUAREZ, M. (ed.). **Fictions in Science**: Philosophical Essays on Modeling and Idealization. London: Routledge, 2009c.
WINSBERG, E. Handshaking Your Way to the Top: Inconsistency and falsification in intertheoretic reduction. **Philosophy of Science**, v. 73, p. 582-594, 2006.
WINSBERG, E. Simulated Experiments: Methodology for a Virtual World. **Philosophy of Science**, v. 70, p. 105-125, 2003.
WINSBERG, E. Simulations, Models, and Theories: Complex Physical Systems and their Representations. **Philosophy of Science**, v. 68, p. S442-S454, 2001.
WINSBERG, E. Sanctioning Models: The Epistemology of Simulation. **Science in Context**, v. 12, n.3, 275-292, 1999.

Sobre a Organizadora

Luana Poliseli: é pós-doutoranda no *Konrad Lorenz Institute for Evolution and Cognition Research* (KLI), Áustria, onde trabalha na intersecção entre filosofia da ciência e estética. Sua pesquisa busca investigar as formas pelas quais as teorias e noções em estética podem auxiliar e melhor o desempenho científico de construção de modelos, compreensão cientifica, além do uso da arte para a comunicação científica. É pesquisadora colaboradora do grupo *Global Epistemologies and Ontologies of Science (Wageningen University & Research - WUR*) e do Instituto Nacional de Ciência e Tecnologia em estudos Inter- e Transdisciplinares em Ecologia e Evolução (INCT-INTREE UFBA). Realizou doutorado em Ensino, Filosofia e História das Ciências (UFBA), com período sanduíche na *University of Amsterdam* (UvA), mestrado e graduação em Ciências Biológicas (UFPB | FAP). Áreas de atuação: filosofia da ciência em prática, história e filosofia da biologia, sociologia do conhecimento, epistemologia científica, epistemologia social, estética. Temas de pesquisa: explicação e compreensão científica; estética e construção de modelos; desacordos acadêmicos; conhecimento ecológico tradicional e acadêmico, inter- e transdisciplinaridade.

Sobre os Tradutores e Revisores

Ana Margarete Barbosa de Freitas: possui graduação em Psicologia pela Universidade Federal da Bahia (2006). Tem experiência na área de Psicologia, com ênfase em Avaliação de Projetos Sociais e Políticas Públicas (metodologia qualitativa), e em magistério do ensino superior, na área de Psicologia da Educação e Educação Emocional. É membro do grupo de pesquisa Mente, Realidade e Conhecimento (UFBA). Tem interesse nas áreas de Filosofia da Mente e da

Linguagem, Epistemologia Analítica, Psicologia e Ciência, Psicologia e Educação. Participou do Programa de Cooperação Acadêmica UFBA/USP realizando estágio de pesquisa na área de Epistemologia e Filosofia da Mente na USP. Possui mestrado em Filosofia Contemporânea (UFBA), cuja pesquisa investigou a natureza da mente e as implicações e limitações do estudo científico dos eventos mentais (Psicologia Científica). Concluiu uma pesquisa de doutorado sobre agência epistêmica e o valor da reflexão na satisfação de objetivos epistêmicos e nos processos de formação de crença no Programa de Pós-Graduação em Filosofia da UFBA. Como parte desta pesquisa, participou do Programa de Doutorado Sanduíche no Exterior na Universidad Nacional Autónoma de México (UNAM) no Instituto de Investigações Filosóficas, com apoio financeiro da CAPES.

Felipe Rocha Lima Santos: Atualmente, é pesquisador pós-doutoral no Programa de Pós-Graduação em Filosofia da UFBA (PNPD). Na UFBA, além de dedicar grande parte do tempo em pesquisa e publicações, também é responsável por disciplinas de Epistemologia na pós-graduação em Filosofia, participação de grupos de pesquisa e organização de eventos a serem realizados nesta universidade. Além disso, é pesquisador associado do CONCEPT - *Cologne Center for Contemporary Epistemology and the Kantian Tradition*, centro de pesquisa em Epistemologia da Universidade de Colônia, Alemanha, sob responsabilidade de Sven Bernecker. Em 2018 publicou seu primeiro livro chamado "Epistemologia e Virtudes Intelectuais: Do Conhecimento ao Entendimento", fruto de sua pesquisa de doutorado. Anteriormente, trabalhou como pesquisador pós-doutoral na *Universidade Nacional Autónoma de México* (UNAM), dedicando seu tempo à pesquisa e ao ensino nos programas de pós-graduação em Filosofia e em Filosofia das Ciências. É Doutor em Filosofia (UFBA, 2015), com pesquisa realizada na área de Epistemologia Contemporânea, sob orientação de Waldomiro J. Silva Filho. Realizou o Doutorado Sanduíche (Programa PDSE - Capes) na *University of California – Irvine* (2013), tendo sido orientado por Sven Bernecker. Em seu mestrado (UFBA, 2010), realizou pesquisa na área de Filosofia da Linguagem, estudando Wittgenstein, sob orientação de Alexandre Machado Noronha.

Jessica Suellen Oliveira Costa: Licenciada em Ciências Biológicas pela União Metropolitana de Educação e Cultura (2018). Atualmente mestranda no PPGECOTAV da Universidade Federal da Bahia. Tem experiência nas áreas de Educação, Ecologia e Zoologia, com ênfase em Comportamento e Cognição Animal,

atuando principalmente nos seguintes temas: plasticidade fenotípica e cognição estendida.

Lúcia Carvalho Neco: Doutoranda em Filosofia no Departamento de Política, Mídia e Filosofia da *La Trobe University*, Austrália. Mestre em Comportamento Animal no Programa de Pós-Graduação em Psicologia Experimental pela Universidade de São Paulo - USP. Graduada em Ciências Biológicas pela Universidade Federal da Bahia - Brasil e Graduanda em Direito pela mesma Universidade. Trabalhou como consultora técnica da empresa Ambiente Sustentável em Salvador, Bahia entre os anos de 2017-2019. Participa do Laboratório de Etologia, Ecologia e Evolução de Insetos Sociais - LEEEIS, USP, Laboratório de Ensino, Filosofia e História da Biologia - LEFHBio, UFBA e do Núcleo de Etologia e Evolução - NuEvo, UFBA. Foi bolsista CAPES do Programa Ciência sem Fronteiras - Graduação Sanduíche na *University of California, Davis* - Estados Unidos colaborando com o *Cultural Evolution Lab* e realizou estágio temporário em *Blumstein Lab na University of California*, *Los Angeles*. Interessada em comportamento animal e filosofia da biologia, especialmente no estudo do comportamento social e evolução, além de legislação ambiental e conservação.

Mariana Vitti Rodrigues: Possui graduação em Licenciatura Plena (2010) e Bacharelado (2012) em Filosofia pela Universidade Estadual Paulista Júlio de Mesquita Filho, mestrado em Filosofia pela mesma universidade (2014) e doutorado em Filosofia da Ciência pela Universidade de Copenhagen (2019). Atualmente é pós-doutoranda em Filosofia da informação e da Tecnologia financiada pela FAPESP (n. 2020/03134-1). Interesse de pesquisa centrado nos conceitos de abdução, descoberta científica, informação, e estilos de raciocínio, bem como nas consequências epistemológicas do extensivo uso de tecnologias de Big Data na prática científica.

Nathalia Cristina Alves Pantaleão Strongen: Doutoranda em Filosofia pela UNICAMP. Mestre em Filosofia pela Unesp - Campus de Marília. Licenciada em Filosofia pela mesma instituição. Membro do Grupo Interdisciplinar "Sistêmica e Auto-Organização" - Cle Unicamp e do Grupo Acadêmico em Estudos Cognitivos da Unesp Marília. Possui experiência em Filosofia da Mente e Ciência Cognitiva, principalmente no ramo de Inteligência Artificial e Modelos Computacionais. Atualmente pesquisa sobre as possibilidades e os limites explicativos da Teoria Computacional da Mente no que diz respeito à capacidade semântica dos indivíduos. Em 2013 foi premiada com o Prêmio de Melhor Trabalho de Filosofia da Mente no VII Colóquio

Internacional de Filosofia da Mente. Em 2015 foi premiada com o 3º Lugar do "Prêmio Marcelo Dascal de Ciência Cognitiva" organizado pela Sociedade Brasileira de Ciência Cognitiva. Editora do Periódico Revista Kínesis de Filosofia da Unesp (qualis B2). Em 2018 realizou estágio de pesquisa na *University of Delaware* sob a supervisão do professor Frederick Adams, e em 2019 realizou visita técnica na Universiteit Antwerpen sob supervisão do professor Erik Myin. Possui apresentações de trabalhos na Argentina, Uruguai e Portugal, além de diversas regiões do Brasil, participando assim, da comunidade acadêmica de língua portuguesa, espanhola e inglesa.

Pedro Bravo de Souza: Está, atualmente, no doutorado em Filosofia pela USP com bolsa processo 2019/10200-3 da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). É mestre, bacharel e licenciado em Filosofia pela UNESP - Marília, além de Bacharel em Relações Internacionais (Uninter - EaD). Realizou graduação-sanduíche e estágio de pesquisa durante o mestrado na *Université Lille 3*, na França. É membro do corpo editorial da Revista Kínesis. Possui especial interesse por temas na intersecção entre Filosofia da Ciência, Epistemologia e Ética. Realiza divulgação filosófica através do projeto *Filosofia a Sério*.

Renata Arruda: Possui graduação e mestrado em Filosofia pela Universidade Federal de Goiás, e doutorado em *Lógica y Filosofía de la Ciencia pela Universidad de Salamanca (*Espanha), durante o qual realizou uma estância de investigação na *Université Paris 1 Panthéon-Sorbonne* (França). É professora efetiva do Curso de Filosofia da Regional Cidade de Goiás, da Universidade Federal de Goiás. Tem experiência nas áreas de Filosofia da Ciência e Lógica, e formação acadêmica nos temas de intersecção entre Filosofia da Ciência e causalidade, condições necessárias e suficientes, problema da indução, e filosofia da medicina e da epidemiologia. Atualmente investiga questões metafísicas e epistemológicas acerca das ciências da saúde, e mais especificamente acerca do papel da manipulação na Medicina.

Tiago Luís Teixeira de Oliveira: Bacharel e Licenciado em Filosofia pela Pontifícia Universidade Católica de Minas Gerais (2005). Mestre (2011) e Doutor em Filosofia pela Universidade Federal de Minas Gerais - UFMG (2017). Fez estágio de pós-doutorado na Universidade do Estado do Rio de Janeiro (UERJ) de agosto de 2019 a julho de 2020. É professor do Ensino Básico Técnico e Tecnológico (EBTT) no Colégio Pedro II (RJ), sendo membro do Núcleo Docente Estruturante (NDE) da Licenciatura em Filosofia e pesquisador do Laboratório de Humanidades

(LabHum) do Campus de Niterói. É também membro do corpo editorial do periódico Investigação Filosófica (ISSN 2179-6742) e participa do Grupo de Pesquisa Investigação Filosófica (UNIFAP). Tem pesquisa nas áreas de Filosofia da Ciência e em Ensino de Filosofia. Interessa-se também por Epistemologia e Metafísica.

Editora
UFPel

DISSERTATIO
FILOSOFIA