



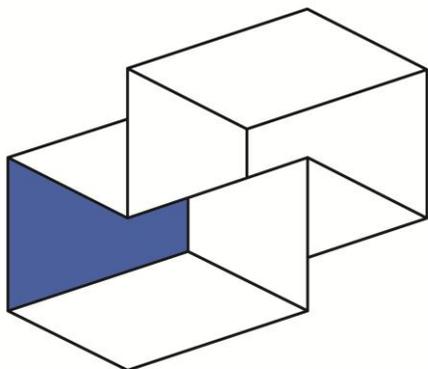
Investigação Filosófica

Revista de Filosofia

ISSN: 2179-6742

Investigação Filosófica, v. 5, n. 2, Ago./Dez., Rio de Janeiro, 2014, 207 p.

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE FILOSOFIA E CIÊNCIAS SOCIAIS
PROGRAMA DE PÓS-GRADUAÇÃO LÓGICA E METAFÍSICA



PPGLM

**Programa de Pós-Graduação
Lógica e Metafísica**

Pós-Graduação em Filosofia - UFRJ

Coordenador

Rodrigo Guerizoli

Vice-Coodenador

Antonio Frederico Saturnino Braga

Revista desenvolvida em parceria com o Programa de Pós-Graduação Lógica e Metafísica (PPGLM) da Universidade Federal do Rio de Janeiro (UFRJ).

INVESTIGAÇÃO FILOSÓFICA
<http://periodicoinvestigacaofilosofica.blogspot.com.br/>
ifilosofica@gmail.com

Editores Responsáveis

Rodrigo Reis Lastra Cid
Luiz Maurício Bentim da Rocha Menezes

Coordenadores Editoriais

Luiz Helvécio Marques Segundo
Mayra Moreira da Costa
Pedro Vasconcelos Junqueira Gomlevsky

Conselho Editorial

Danillo de Jesus Ferreira Leite
Guilherme da Costa Assunção Cecílio
Leandro Shigueo Araujo
Luis Fernando Munaretti da Rosa
Luiz Helvécio Marques Segundo
Luiz Maurício Bentim da Rocha Menezes
Mário Augusto Queiroz Carvalho
Mayra Moreira da Costa
Paula Akemy Araujo
Pedro Vasconcelos Junqueira Gomlevsky
Renata Ramos da Silva
Rodrigo Alexandre de Figueiredo
Rodrigo Reis Lastra Cid
Sagid Salles Ferreira
Tiago Luís Teixeira de Oliveira

Conselho Consultivo

Alexandre Meyer Luz
Alexandre Noronha Machado
Carlos Eduardo Evangelisti Mauro
Desidério Orlando Figueiredo Murcho
Guido Imaguire
Mário Nogueira de Oliveira
Michel Ghins
Roberto Horácio de Sá Pereira
Rodrigo Guerizoli Teixeira
Rogério Passos Severo
Sérgio Ricardo Neves de Miranda
Ulysses Pinheiro

Equipe Técnica

Logotipo: Thiago Reis
Revisor: Fábio Salgado de Carvalho
Tradutor: Giuliano Tadeu Nunes Pietoso

INVESTIGAÇÃO FILOSÓFICA

Revista de Filosofia Semestral
Volume 5, número 2, 2014, 207p.
Publicação digital
ISSN: 2179-6742

1. Filosofia – Periódicos. 2. Universidade Federal do Rio de Janeiro. Instituto de Filosofia e Ciências Sociais. Programa de Pós-Graduação Lógica e Metafísica. 3. Blog Investigação Filosófica.

Sumário / Contents

Editorial	01
------------------------	----

Artigos/Articles

Do não cognitivismo ao cognitivismo dos enunciados éticos - a mudança nas noções de fato e de objetividade em Hilary Putnam

<i>Alexandre de Freitas de Mello Junior</i>	02
---	----

O medo e a origem do Estado em Thomas Hobbes

<i>Gerson Vasconcelos Luz</i>	29
-------------------------------------	----

Da predileção à exigência: a concepção kierkegaardiana do amor enquanto dever

<i>José C. Lopes Marques & José Roberto Gomes da Costa</i>	36
--	----

Traduções/Translations

O trabalho recente sobre a controvérsia internismo-externismo

Laurence Bonjour

<i>Tradução de Luiz Helvécio Marques Segundo</i>	54
--	----

Por que não há mais progresso na filosofia?

David J. Chalmers

<i>Tradução de Gregory Gaboardi</i>	71
---	----

Lógica modal

James Garson

<i>Tradução de Heitor Hedler Siqueira Campos</i>	102
--	-----

Metodologia: os elementos da filosofia da ciência

David Papineau

<i>Tradução de Luiz Helvécio Marques Segundo</i>	131
--	-----

Investigação Editorial

É com muito prazer que damos início ao volume 5, número 2 da revista *Investigação Filosófica*. Nessa edição temos um artigo sobre a mudança nas noções de fato e de objetividade em Hilary Putnam de Alexandre de Freitas de Mello Junior. O segundo artigo trata sobre o medo como uma paixão fundamental para a origem do Estado em Thomas Hobbes de Gerson Vasconcelos Luz. O terceiro artigo pretende analisar o conceito kierkegaardiano de amor a partir de sua relação com o dever e da compreensão de *próximo* enquanto *reduplicação* José C. Lopes Marques e José Roberto Gomes da Costa. A revista também está composta de quatro traduções, a primeira é “O trabalho recente sobre a controvérsia internismo-externismo” de Laurence BonJour; a segunda é “Por que não há mais progresso na filosofia?” de David J. Chalmers; a terceira é “Lógica modal” de James Garson; e a quarta é “Metodologia: os elementos da filosofia da ciência” de David Papineau.

Desejamos a todos uma boa e agradável leitura filosófica.

Rodrigo Cid
Luiz Maurício Menezes

DO NÃO COGNITIVISMO AO COGNITIVISMO DOS ENUNCIADOS ÉTICOS - A MUDANÇA NAS NOÇÕES DE FATO E DE OBJETIVIDADE EM HILARY PUTNAM

Alexandre de Freitas de Mello Junior¹

RESUMO: Este artigo será dividido em duas partes. Na parte inicial, procuraremos mostrar que o positivismo lógico é tributário de uma noção de fato herdada do empirismo clássico. Nosso interesse é revelar a consolidação dessa dicotomia², que em muitos aspectos é análoga à dicotomia entre fato e valor. Ambos os binômios conceituais, que de início eram simples distinções, tornaram-se dicotomias por meio de uma inflação dos seus conceitos. Na segunda parte, apresentaremos a dicotomia fato/valor, mostrando o périplo do argumento putnamiano, que irá nos revelar o imbricamento entre as noções de fato e valor. Assim, é interessante apresentar o raciocínio que revelará a Putnam a constatação da fratura interna de uma dicotomia (analítico/sintético), impetrado por Quine³, para, posteriormente, chegarmos à ideia principal do solapamento da dicotomia fato e valor. Depois disso feito, poderemos considerar a suposição da objetividade (de alguns) dos valores (éticos).

PALAVRAS-CHAVES: Empirismo; fato; valor; dicotomia; objetividade; ética.

Abstract: This article will be divided into two parts. In the first part, we aim to show that logical positivism is a notion of tax actually inherited from classical empiricism. Our interest is to reveal the consolidation of this dichotomy, which in many ways is analogous to the dichotomy between fact and value. Both conceptual dichotomies, which initially were simple distinctions became dichotomies through an inflation of its concepts. In the second part, we present the fact / value dichotomy, showing the journey of putnamiano argument, which will reveal the interweaving notions of fact and value. Thus, it is interesting to present the reasoning that reveal the Putnam examining the internal fracture of a dichotomy (analytic / synthetic), filed by Quine, to then get to the main idea of undermining the dichotomy between fact and value. Once that is done, we may consider the assumption of objectivity (some) values (ethical).

Keywords: Empiricism; fact; value; dichotomy; objectivity; ethics.

O positivismo lógico, que Putnam faz uso para iniciar a sua crítica, apresenta uma versão tripartite dos juízos, que compreende as duas dicotomias aqui aventadas: os juízos sintéticos, que são verificáveis empiricamente; os juízos analíticos, que podem ter um valor de verdade a si atribuído simplesmente em virtude do significado dos termos

¹ Mestrando da Universidade de Brasília.

² Não temos o intuito de fazer uma análise crítica da distinção entre o analítico e sintético. Nosso objetivo, ao analisar essa distinção, é apresentar a formação de uma dicotomia, ao invés de expor os motivos pelos quais muitos enunciados não podem ser classificados como analítico ou sintético, de maneira estanque.

³ Quine é o autor do artigo *Dois Dogmas do Empirismo*, no qual ele apresenta a questão de um dos dogmas, a divisão entre as verdades analíticas e as verdades sintéticas estar mal fundamentada. Putnam não se delongará na argumentação quineana sobre a dicotomia analítico/sintético, pois não é o argumento de Quine em si que está no centro de sua análise. O que nos interessa aqui é mostrar que a distinção entre o analítico e o sintético não leva a uma dicotomia, bem como no caso da distinção entre os juízos de fato e os juízos de valor.

constantes nessa classe de enunciados, prescindindo, desse modo, de verificações empíricas; e, por fim, os juízos éticos, metafísicos e estéticos, que são aqueles juízos não cognitivos, ou seja, juízos destituídos de significado.

A primeira coisa a ser dita parte do ponto da caracterização do âmbito ao qual a dicotomia analítico/sintético pode ser aplicada. Segundo a distinção entre analítico e sintético, os enunciados da linguagem que são cognitivamente relevantes poderão ser verdadeiros ou falsos, consoante os critérios da classe aos quais se referem. Assim, serão verdadeiros ou falsos, do ponto de vista da classe de enunciados analíticos, em razão das regras lógicas do enunciado, dessa forma prescindindo da experiência; e serão verdadeiros ou falsos, do ponto de vista de enunciados sintéticos, em razão dos testes experimentais, passíveis de observação.

A linguagem cotidiana, ao contrário da linguagem cognitivamente relevante, é limitadora, pois nela recorrentemente há uma possibilidade de confusão, ou melhor, uma imprecisão habitual nos seus termos, na ausência de uma forma, e na possibilidade de constatar a verdade ou falsidade de seus enunciados (por meio verificacional). Quando a linguagem torna-se cognitivamente relevante, resulta que os enunciados serão verdadeiros ou falsos e, por isso mesmo, ou passíveis de teste segundo o modelo observacional, ou demonstráveis segundo suas regras da linguagem constituída⁴.

Da mesma forma que Putnam nos demonstrará com relação à dicotomia fato/valor, ele nos ensina que a dicotomia analítico/sintético, ora abordada, assumiu, também, um tom onipresente (PUTNAM, 2008a, p. 23-24). Isso significa que todos os juízos recairão sob essa classificação que tem pretensão universal, no tocante a juízos com significados. Assim, todo juízo cognitivamente relevante é ou analítico, ou sintético. Os demais juízos fazem parte do conjunto de juízos cognitivamente não relevantes.

Os positivistas lógicos fizeram a totalidade dos juízos cognitivamente relevantes repousarem nas categorias de sintético ou analítico. Desse modo, dada a universalidade da aplicação da dicotomia metafísica, todos os juízos (cognitivamente relevantes) estariam inseridos em uma ou outra categoria. Portanto, os enunciados científicos deveriam ser classificados segundo esse estatuto. A física teórica, por exemplo, estaria no conjunto de enunciados que deveriam ser classificados dessa maneira para que

⁴ O caso de certos enunciados serem demonstráveis em virtude do significado dos termos constantes no enunciado.

pudessem receber a chancela de racional. Isso quer dizer que tais enunciados só seriam significativos caso pudessem ser formulados de modo a gerar uma resposta que indicará se eles são analíticos ou sintéticos, distintamente (2008a, p. 25).

Quine, ao contrário, empreende uma sequência que vai em direção à dissolução completa não só da dicotomia, mas também da distinção⁵ entre o analítico e o sintético. Se o sintético e o analítico fossem exemplificados de forma metafórica pelas cores “branco” e “preto”, estariam, seguindo a metáfora, mais próximo da cor “cinza”, o que simbolizaria o caráter não distinguível de ambos os enunciados⁶:

O saber de nossos pais é um tecido de sentenças. Em nossas mãos, ele se desenvolve e transforma, por meio de nossas próprias revisões e adições mais ou menos arbitrárias e deliberadas, mais ou menos diretamente ocasionadas pelo estímulo contínuo de nossos órgãos sensoriais. É um saber cinza pálido, preto para o fato e branco para a convenção. Mas não encontrei razões substanciais para concluir que exista nele algum fio totalmente preto ou totalmente branco (QUINE, 1963, p. 406 *apud* PUTNAM, 2008a, pp. 25-26).

Segundo Putnam, esta ideia de que não há qualquer distinção entre as duas classes de juízos não deve ser levada tão longe, uma vez que ele próprio aceita que certos juízos são ou analíticos ou sintéticos. Porém, o que ele retém da crítica de Quine e, conseqüentemente, aceita como uma sugestão reproduzível no caso da dicotomia fato/valor é a observação de que determinados juízos cognitivos não recaem nos tipos de classes ‘analítico’ e ‘sintético’. (PUTNAM, 2008a, p. 26). Igualmente, Putnam faz uma importante consideração que merece destaque, a saber, que se temos que determinado tipo de juízo não pode ser classificado como sendo verdadeiro

⁵ Em *Os dois dogmas do empirismo* ele rejeita a distinção entre os chamados “juízos analíticos” e os “juízos sintéticos”: “Minha proposta atual é que é um contrassenso, e a base para muitos outros contrassensos, falar de um componente linguístico e de um componente factual na verdade de qualquer enunciado individual. Tomada coletivamente, a ciência tem sua dupla dependência da linguagem e da experiência, mas essa dualidade não é significativamente delineável nos enunciados da ciência tomados um a um. (QUINE, 2011, p. 66). Embora pareça, ao contrário do que pensa Putnam, que Quine rejeita a distinção absoluta, que tem como consequência a dicotomia. (ver QUINE, 2011, p. 71). Não obstante, a interpretação de Putnam está corroborada pelo texto supracitado.

⁶ Putnam lembra que Quine, ao lançar uma crítica ao positivismo lógico em seu artigo intitulado *Carnap and Logical Truth*, dirige-se, em primeiro lugar, para a distinção clara e estanque que os positivistas lógicos fazem com respeito aos aspectos inter-teóricos, como a elegância (contraparte convencional) e o teste experimental dos enunciados hipotéticos (o que os positivistas chamam de ‘fato’). Segundo ele, um elemento de convenção é uma característica que poderá, quando adotada anteriormente, ser posteriormente submetida a testes. A citação de Quine que Putnam recupera escapa ao âmbito teórico específico à linguagem científica, mas não à linguagem que os positivistas lógicos considerariam factual, ou seja, falseável por poderem ser submetidas a testes baseados em experiências concretas. Desse modo, recaem tanto os enunciados científicos da física, que são sintéticos, quanto o “saber de nossos pais”, no âmbito do factual, por serem ambos passíveis de verificações a posteriori.

simplesmente em razão das regras lógicas (analítico), ele não deve recair como sendo um enunciado sobre questões de fato ou enunciados do tipo sintético.

Nessa esteira, Putnam declara que pode ser feita uma crítica ao caráter compulsório da dicotomia analítico/sintético, em que ambos os lados da dicotomia são naturalizados, ou seja, são consideradas as únicas classes básicas de enunciados com sentido. Dessa forma, insere-se nessas duas classes básicas de enunciados uma característica essencial compartilhada, qual seja, a de que juízos não analíticos são complementares aos juízos analíticos, e suas propriedades intrínsecas alternam-se sempre que um juízo passa de uma classe para outra.

Outrossim, sempre que os juízos não forem verdadeiros em razão das regras da linguagem lógica, serão juízos com a propriedade de serem “descrições de fato, em que o modelo original de um fato é o tipo de fato empírico que se pode *figurar*” (2008, p.27). A dificuldade aqui verificada é o fato de não se perceber outros tipos de enunciados que não sejam trivialmente verdadeiros, ou seja, que existe uma categoria de enunciados cognitivamente relevantes que escape à categoria dos juízos não analíticos típicos.

Como será possível ver, a dicotomia analítico/sintético tem muito a revelar no tocante à dicotomia fato/valor, aqui privilegiada, pois ambas possuem desenvolvimentos semelhantes. Putnam nos faz atentar para um ponto em comum, dos mais centrais, para o qual estas duas dicotomias convergem, e que as fundamenta em sua história conceptual, qual seja, a Lei de Hume⁷. Segundo esta lei, grosso modo, não podemos derivar um ‘deve’ a partir de um ‘é’⁸. Isso significa que nenhuma noção ética pode ser inferida de uma questão de fato, o que impossibilitaria de se ter um juízo moral ancorado em um fato da realidade.

⁷ Putnam parece seguir a interpretação quineana de que a “distinção kantiana entre verdades analíticas e sintéticas já se prefigurava na distinção de Hume entre relações de ideias e questões de fato”. (QUINE, 2011, p. 37) Do mesmo modo, a dicotomia fato/valor encontra nessa distinção de Hume o ponto de partida de sua formação.

⁸ O trecho do texto de Hume que se refere abertamente à famosa Lei de Hume é o seguinte: “Em todo sistema de moral que até hoje encontrei, sempre notei que o autor segue durante algum tempo o modo comum de raciocinar, estabelecendo a existência de Deus, ou fazendo observações a respeito dos assuntos humanos, quando, de repente, surpreendo-me ao ver que, em vez das cópulas proposicionais usuais, como é e não é, não encontro uma só proposição que não esteja conectada a outra por um deve ou não deve. Essa mudança é imperceptível, porém da maior importância. Pois como esse deve ou não deve expressa uma nova relação ou afirmação, esta precisaria ser notada e explicada; ao mesmo tempo, seria preciso que se desse uma razão para algo que parece totalmente inconcebível, ou seja, como essa nova relação pode ser deduzida de outras inteiramente diferentes”. HUME, David. Tratado da Natureza Humana. Tradução de Débora Danowski. Livro III, Parte I, Seção II. São Paulo, Editora UNESP, 2000, p. 509

Mas o fundamental da Lei de Hume, para Putnam é outro. Para o filósofo, o tema da inderivabilidade não demonstrável exposta na Lei de Hume não tem seu cerne simplesmente naquilo que diz respeito à questão inferencial, donde não poderíamos, claramente, apresentar como teorema a passagem lógica do ‘é’ para o ‘deve’. Segundo ele:

Se a afirmação fosse simplesmente acerca da *forma* de certas inferências, ela proibiria qualquer um de inferir “deve-se fazer *x* em tais e tais circunstâncias” a partir de “em tais e tais circunstâncias, é bom fazer *x* e, nessas circunstâncias, é mau deixar de fazer *x*”. Por certo, muitos filósofos responderiam a esse exemplo dizendo que ele não entra em conflito com o *dictum* de Hume, porque é um caso de inferência de um “deve” a partir de outro “deve”. Mas esse é o meu ponto. A habilidade deles de reconhecer enunciados tais como “em tais e tais circunstâncias, é bom fazer *x* e, nessas circunstâncias, é mau deixar de fazer *x*” como um caso de “deve” não resulta de alguma característica da forma do enunciado mas, antes, de um entendimento do seu conteúdo (PUTNAM, 2008a, pp. 28-29)

Esta consideração é reveladora de um aspecto metafísico componente da percepção do problema do ponto de vista humeano, pois toma como motor da Lei uma constatação dos enunciados, que incluem o ‘deve’, não do ponto de vista lógico, demonstrando, assim, que a questão é muito mais profunda do que uma simples consideração formal. Essa inclusão da centralidade do conteúdo, ao invés do privilégio do modo inferencial, que objetaria que a passagem do ‘é’ para o ‘deve’ acarreta a invalidade do procedimento lógico, nos conduz a uma perspectivação do problema que é, quando colocado assim, dimensionado de outra maneira. Destarte, a centralidade do conteúdo exigirá que tenhamos em consideração aquilo faz, para Hume, que um enunciado seja um enunciado de ‘é’, pois assim entenderemos o que ele quer dizer com a noção de ‘fato’.

A observação que Putnam faz nos lembra que Hume considera, em sua teoria da mente, que os conceitos são espécies de ideias, e estas são figurativas. Desse modo, as ideias humanas somente representam “questões de fato” quando a elas se *assemelham* (2008a, p.29), o que significa que as figuras são formadas a partir de nossa experiência sensível, que é, também, o lastro no qual estará amparado a noção última de fato.

Com isso, pode-se perceber que Hume não interpõe a sua “Lei” meramente à impossibilidade da passagem de enunciados factuais a enunciados valorativos, mas que, em razão da maneira como as ideias recomendam a noção de fato, não há, no tocante às

virtudes, realmente nenhuma “questão de fato”⁹. O motivo pelo qual isso não ocorre é porque, com relação às virtudes, não temos como formar uma imagem figurável das propriedades de uma determinada virtude hipotética, ao passo que poderíamos, sem a menor dúvida, fazer figurar as propriedades de uma flor qualquer, por exemplo.

Entre o analítico e o sintético, ambos os lados da dicotomia assinalam uma série de propriedades ou características que são, entre si, distintas. O mesmo ocorre com a dicotomia fato/valor, pois, segundo o entendimento de Putnam, os juízos que envolvem a noção de “deve” são a parte do lado valorativo ao qual está vedada uma relação com os juízos de fato. Algumas palavras também se refeririam a essa classe de juízos, tais como “correto”, “errado”, “bom”, “mau”, quando usados, evidentemente, em suas acepções éticas¹⁰ (2008a, p. 31). Essa distinção humeana, quando inflada, ou seja, levada ao extremo da distinção, impõe que não teremos, e não poderemos ter juízos de fato que se refiram à ética. Desse modo, o argumento pode ser colocado da seguinte maneira:

(1) o que nos possibilita chamar algo de factual, ou seja, ser um enunciado de fato, é conseguirmos percebê-lo por meio do nosso aparato sensório (ter uma impressão), sendo, somente assim, passível de representação.

(2) são “objetos” somente aquelas entidades que podem ser capturadas por meio do aparato sensório.

(2.1) ser capturado por meio do aparato sensório confere “objetividade” à entidade assim capturada.

(3) se somente podemos atribuir objetividade aos objetos, então todos aqueles enunciados que não correspondam a enunciados factuais não são objetivos. 1 e 2.

(4) Os enunciados de “deve” não podem ser capturados pelo aparato sensório.

Portanto, os enunciados de “deve” não são enunciados objetivos. 3 e 4.

O argumento, que é uma reconstrução que remete a Hume, apesar de válido, só pode estar correto se suas premissas forem verdadeiras. Desse modo, a tarefa de Putnam

⁹ Para corroborar isso, segue o texto do próprio Hume: “Se o pensamento e o entendimento sozinhos fossem capazes de fixar os limites do certo e do errado, a qualidade de virtuoso ou vicioso teria de estar em algumas relações de objetos, ou então ser uma questão de fato, descoberta por nosso raciocínio.” (HUME, 2009, pp. 502-503).

¹⁰ Quando falamos que um sapato é bom, ou que uma operação aritmética está errada, não usamos “bom” e “errada” em acepções éticas. Obviamente, a classe resultante da coleção de enunciados de “deve”, e seus correlatos, são aqueles usados na acepção ética.

será reconfigurar a noção de fato. Ele usará, como estratégia, a apresentação do solapamento da dicotomia fato e valor.

Segundo a divisão apresentada por Putnam, há uma necessidade no tocante à inclusão dos enunciados em uma ou outra classe de enunciados analíticos ou sintéticos. Como já foi dito, se um enunciado (cognitivamente relevante) não for analítico, necessariamente esse enunciado (cognitivamente relevante) será um enunciado não analítico. Forçosamente, em razão da dicotomia, todo enunciado não analítico é um enunciado sintético. De modo análogo, todo enunciado não ético (correlacionado a “deve”) é, necessariamente, um enunciado de fato. (2008, p. 32).

Essa fórmula dicotômica repercute de maneira clara no modo como os objetos são configurados, pois a distinção, inflada até tornar-se dicotomia, subsidia o formato metafísico assumido, no qual ou se é um enunciado de fato, ou se é um enunciado de valor. Não há, aqui, lugar para a classe que Quine, recuperado por Putnam, chama de “cinza”, mas tão somente para as classes “preto” e “branco”.

A respeito do que temos falado, cabe lembrar que os “valores” em questão sempre são tratados, dentro da história da dicotomia, como valores éticos, omitindo-se, habitualmente, qualquer emergência de valores que não sejam éticos¹¹. Putnam alega que “Desde Hume o fato de que existam muitos tipos de juízos de valor que não são em si mesmos de uma variedade ética (ou moral) tende a ser colocado de lado nas discussões filosóficas da relação entre os (chamados) valores e os (chamados) fatos” (2008a, p. 34).

Assim, os valores não éticos não têm sido considerados, e a dicotomia, inflada dessa maneira, omite uma importante categoria que poderia lançar nova luz ao problema. É isso que Putnam também resgatará dentro de sua estratégia argumentativa.

Putnam faz uma importante consideração sobre a natureza da dicotomia. Sua reflexão é que se tivermos o binômio fato/valor considerado somente no âmbito de uma importante distinção conceitual filosófica, não se seguirá dela nenhuma metafísica, do mesmo modo que não se segue nenhuma metafísica do fato de distinguir-se entre

¹¹ Temos, também, como exemplo de valores que não são nem éticos ou não éticos os valores estéticos. Entretanto, estes padecem do mesmo problema considerado na dicotomia, pois não são factuais e cognitivos. Putnam deixa isso claro na parte II do livro “O Realismo de Rosto Humano”, de 1999, intitulada “Ética e Estética”.

enunciados relativos à química e enunciados que não são relativos à química, por exemplo (2008a, p.35).

A importância da distinção estaria posta pela obviedade de que fato e valor não são sinônimos, e por isso mantêm peculiaridades e acepções específicas com respeito ao seu significado. Porém, ao traçar a distinção, não há nenhuma necessidade de uma decorrência metafísica ocasionada pelo que temos chamado de “dicotomia”. Afinal, distinguir a cor “azul” da cor “amarelo” não é o mesmo que afirmar que elas são dicotômicas, por exemplo.

Na próxima seção, passaremos à análise da noção de fato que os positivistas lógicos adotaram, para lançarmos uma crítica com respeito aos pressupostos nos quais ele mesmo se ampara e fundamenta. Feito isso, poderemos começar a apresentação do argumento que irá culminar no colapso da dicotomia entre fato e valor.

1. 1 A Crítica de Putnam à Noção de Fato do Positivismo Lógico

Pelo que vimos até aqui, podemos afirmar que a dicotomia entre fato e valor repousa sobre a noção de fato. Putnam nos explica que “a noção humeana de um ‘fato’ é simplesmente aquilo do qual pode existir uma ‘impressão’ sensível” (2008a, p. 38). Como já foi aventado acima, essa noção de fato é constitutiva da consolidação da dicotomia entre fato e valor e, dito de outro modo, isso pode ser asseverado por não possuímos, no nosso aparato sensorial, um sentido que consiga perceber uma virtude da mesma maneira que conseguimos perceber uma árvore. Isso limita nosso campo de possibilidade a afirmações factuais. Não temos um órgão sensorial que nos possibilite ter uma impressão do “bem”, ou do “correto”. Portanto, do ponto de vista da filosofia de Hume, estaríamos fadados a ter que manter enunciados éticos separados dos enunciados de fato.

Essa noção de fato foi muito influente na tradição analítica inaugurada pelo Círculo de Viena¹². Essa tradição estava temporalmente distante do empirismo clássico, porém, algumas de suas noções iniciais estavam muito de acordo com ele. Podemos

¹² O primeiro capítulo desse trabalho será dedicado a investigar as decorrências dos pressupostos adotados pelo Círculo de Viena para a ética.

dizer, juntamente com Putnam, que a noção de fato adotada inicialmente pelos filósofos do Círculo de Viena era bastante semelhante ao que Hume declara que ela é¹³.

Putnam nos diz que “a visão positivista lógica original era de que um ‘fato’ era algo que podia ser certificado por mera observação e até mesmo por um mero relato de uma experiência sensorial.” (2008a, p. 39). Assim, os positivistas esperavam poder fornecer ou encontrar um método de verificação que pudesse estabelecer a justificação racional, e esse método, ou métodos, deveriam dar conta de verificar as pretensões cognitivas lançadas pelas diversas classes de enunciados (PUTNAM, 1992, p. 141).

Esses métodos de justificação racional, almejados pelos positivistas lógicos, que encerram a verificação (ou, abstratamente, a demonstração¹⁴), procurarão, assim, arrolar a forma por meio da qual a testabilidade, e a própria descrição dos métodos de justificação racional, poderá dar conta do significado dos enunciados. Essa é uma concepção criteriosa da racionalidade, “segundo a qual existem normas institucionalizadas que definem o que é e o que não é aceitável racionalmente” (1992, p. 146). É intrínseco à concepção criteriosa da racionalidade que esses critérios tenham um caráter público, de forma que deles possam fazer uso qualquer indivíduo que saiba deles dispor.

Desse modo, os enunciados que não puderem ser submetidos ao método serão declarados carentes de significado. Destarte, se um enunciado não puder ser comprovado pelos meios colecionados, quais sejam, métodos matemáticos, métodos lógicos, ou métodos científicos¹⁵, seguir-se-á que não haverá significado cognitivo algum que lhe seja relevante (1992, pp. 141-142). Esse crivo pelo qual devem passar os enunciados com pretensões cognitivas deve ser aplicado a todas as espécies de enunciados.

¹³ Embora os positivistas lógicos tivessem abandonado noções humeanas como “ideia” e “impressão”, para citarmos dois exemplos.

¹⁴ Trata-se de demonstrações lógicas. No caso do critério de que os enunciados cognitivamente relevantes devam ser passíveis de verificação e testes, estamos a falar de enunciados sintéticos; no caso daqueles que são verdadeiros ou falsos em razão das regras lógicas e significado dos termos, estamos a falar dos enunciados analíticos.

¹⁵ É importante notar que até a publicação da obra de Carnap intitulada *Logical Foundations of Probability* (1950), os positivistas lógicos supunham haver, segundo Putnam, “um método formal (lógica indutiva), e que o trabalho continuado podia resultar numa enunciação explícita deste método, uma formalização da lógica indutiva comparável à formalização da lógica dedutiva que fora realizada partindo da obra de Frege 1879” (PUTNAM, 1992, pp. 233-234). As características dessa convicção serão estudadas no capítulo 1 deste trabalho, quando da descrição geral do positivismo lógico.

Entretanto, Putnam afirma, como foi supracitado, que os positivistas, ao fim e ao cabo, possuíam uma noção de fato que estava profundamente ligada à verificação pela via sensorial, ou, dito de outro modo, “que um enunciado significativo deve ser *conclusivamente verificável* por confrontação com a experiência direta” (PUTNAM, 2008a, p.39). Isso não se ajusta, em muitos casos, à adoção da via científica como critério, porque o método científico, várias vezes, postula a existência de entidades que não podem ser percebidas pelos nossos sentidos, como por exemplo, os átomos. Por isso, Putnam nos explica que as mudanças da ciência, na primeira metade do século XX, foram centrais para as sucessivas mudanças por parte dos positivistas lógicos, que objetivavam salvar a noção de fato.

Apesar das mudanças científicas importantes que ocorriam na época, como por exemplo, a física quântica, a teoria da relatividade, as primeiras descrições da estrutura interna do átomo e, apesar de os positivistas lógicos não terem ficado impassíveis diante disso, a noção de fato assumida pelo positivismo lógico permaneceu intacta. Putnam explica-nos que os positivistas lógicos estiveram próximos da noção humeana de fato até por volta de 1936¹⁶. Nesse ano, Carnap flexibiliza o pressuposto “de que todos os predicados factuais devem ser definíveis em termos de observação” (2008a, p. 39).

A mudança é importante, mas a posição do positivismo lógico permanece inalterada em sua parte mais central, uma vez que para um enunciado ser considerado cognitivamente relevante, no que toca ao seu significado, fazia-se necessário que esse enunciado pudesse expressar-se por meio da linguagem científica; e que os predicados constantes na parte factual do jargão científico deveriam apresentar pelo menos uma das duas condições: serem termos coligidos por observação; ou serem redutíveis a termos de observação (2008a, p. 40).

O que podemos afirmar, por meio dessas considerações, é que com o critério de justificação racional assumido pelo positivismo lógico nessa época, não poderíamos dizer que certas entidades postuladas pela ciência existiam *realmente*. Desse modo, em função dessa adoção de critérios de aceitabilidade racional de enunciados, os positivistas lógicos possuíam um universo do discurso tal que não podiam sequer atribuir um caráter cognitivo às várias das descobertas científicas. Essas eram

¹⁶ Putnam usa como marco do início da sutil flexibilização da noção de fato as alterações nos critérios de verificação constantes no artigo de Carnap intitulado *Testability and Meaning* (1936-7), publicado originalmente na revista *Philosophy of Science*.

consideradas um modo de falar acerca do mundo natural, mas que não correspondiam *realmente* a ele.

Entretanto, o critério de significação cognitiva, admitido pelo positivismo lógico para resolver esse problema é que “na medida em que o sistema como um todo permite prever nossas experiências, com maior sucesso do que poderíamos sem ele, tais predicados teóricos devem ser aceitos como ‘empiricamente significativos’¹⁷” (2008a, p. 48), o que provoca uma importante alteração na noção de fato positivista.

Putnam afirma que a linguagem, para os positivistas lógicos, caso quisesse ser cognitivamente significativa, deveria se assemelhar à linguagem que temos na física (2008a, p. 43). Porém, o próprio positivismo lógico mantém pressupostos que não se assemelham a essa linguagem. Putnam afirma que “se for verdade que apenas podem ser aceitáveis racionalmente os enunciados verificáveis criteriosamente, esse mesmo enunciado não pode ser verificado criteriosamente” (PUTNAM, 1992, p. 147).

Portanto, o próprio pressuposto do positivismo lógico, no que diz respeito às condições de verificabilidade e aceitabilidade racional, o derroga, visto que ele não possui os pré-requisitos suficientes e necessários para que faça parte da classe de

¹⁷ Posteriormente, Carnap considerou que entidades físicas, tais como os elétrons, prótons, não são entidades que formam as teorias a partir de simples definições, mas são consideradas entidades primitivas. Segundo Putnam, essa reconsideração provocou uma revisão profunda nos critérios acima mencionados, haja vista que ao serem tomadas como primitivas, certas entidades não precisariam mais ser submetidas aos critérios. Antes, se algum termo científico abstrato pudesse assegurar-nos maior poder preditivo em uma teoria, deveríamos, assim, tomá-lo como “empiricamente significativo”. (PUTNAM, 2008a, p. 41). E prever, para os positivistas lógicos, significa “deduzir sentenças de observação a partir de uma teoria” (2008a, p. 46). No caso presente, essas considerações de Putnam interessam-nos em razão da noção de fato do empirismo lógico. Embora Putnam esteja convencido de que “no critério positivista lógico revisado do significado cognitivo, é o sistema dos enunciados científicos como um todo que tem ‘conteúdo factual’” (2008, p. 41). Isso pode ser afirmado pelo que nós já expomos anteriormente. Dentro de uma teoria científica, o conjunto de enunciados articulados proporciona a capacidade de, por exemplo, prever-se um evento. Ora, se esse conjunto de enunciados, enunciados esses considerados de modo articulado, nos garante que um evento pode ocorrer, devemos concluir que o conjunto de enunciados é significativo do ponto de vista empírico. Mas essa conclusão é válida para o conjunto de enunciados, na perspectiva inter-teórica. Na perspectiva de um enunciado, considerado individualmente, a conclusão muda. Se tivermos, dentro do conjunto de enunciados de uma teoria que garante previsibilidade, um enunciado que postula a existência de um termo abstrato, poderíamos entender o conjunto de enunciados como cognitivamente relevante, pelos motivos já expostos, mas não poderíamos fazer o mesmo com o enunciado individual. Isso deve ao fato de que, tomado individualmente, um enunciado que postula uma entidade abstrata pode estar destituída de meios de confirmação (como, por exemplo, a previsibilidade empiricamente significativa prestada pelo conjunto de enunciados) justamente em razão do seu caráter. Assim, um enunciado individual que postulasse uma entidade abstrata somente seria significativa quando interpretada no interior de uma teoria formada por enunciados empiricamente relevantes. Individualmente, um enunciado de tal espécie não teria relevância cognitiva, ou melhor, não possuiria significado. Por isso, Putnam afirma que a visão geral do conceito de fato do empirismo lógico continuou, no fundo, apesar da revisão do conceito, muito influenciada pelo empirismo clássico. Os enunciados individuais eram aceitos na linguagem científica “como meros artificios para derivar as sentenças que enunciam realmente os fatos empíricos, a saber, as sentenças de observação” (2008a, p. 41).

enunciados que ele mesmo afirma serem aqueles cognitivamente relevantes. É, desse modo, autorrefutante, apesar de sua alteração na noção de fato, pois “o próprio critério dos positivistas lógicos não é (a) analítico, nem (b) testável empiricamente” (1992, p. 142).

Nos dias atuais, embora o positivismo lógico não esteja mais em voga, grande parte dos filósofos analíticos se coaduna à ideia de que a linguagem cognitivamente significativa deve seguir o modelo de uma linguagem física¹⁸. Assim, por exemplo, é muito comum exigir-se na abordagem relativa a “termos psicológicos da linguagem ordinária”, (2008a, p. 44-45) que esses se refiram, necessariamente, a estados cerebrais. Assim, a tristeza, como exemplo de um “termo psicológico da linguagem ordinária”, deveria referir-se a um estado cerebral que lhe corresponda. Esse estado cerebral seria, destarte, a causa do estado psicológico expressa pelo termo em questão.

Diante de tal condicionamento relativo à linguagem cognitivamente relevante, é de se surpreender que possamos fazer qualquer tentativa de prestar aos valores um *status* de significativos. Para Putnam, a linguagem que o positivismo adota se baseia em uma adoção da noção de fato que é bastante limitada e, por isso, é incapaz de perceber que a distinção entre fato e valor não é uma dicotomia.

A dicotomia fato/valor dos positivistas lógicos foi defendida com base em um quadro cientificista estreito do que poderia ser um fato, assim como o ancestral humeano dessa distinção foi defendido com base em uma psicologia empirista estreita das ideias e das impressões (PUTNAM, 2008a, p. 45).

Ante o acima exposto, podemos colocar a crítica ao positivismo lógico e antecipar o colapso da dicotomia fato e valor nos seguintes termos: (1) a linguagem científica, preconizada pelos positivistas lógicos, é pobre, e isso deixa a maior parte do universo discursivo fora do conjunto de termos com significado. E, segundo a crítica de Putnam, pressupostos centrais do próprio positivismo lógico não podem submeter-se com sucesso à concepção criteriosa de racionalidade que lhe é própria, e que exige critérios públicos de verificabilidade que estão postos em sua compreensão da linguagem científica (2) certos termos descritivos do nosso vocabulário avaliativo implicam uma consideração aos fatos, ou seja, há, na concepção de Putnam, termos

¹⁸ Essa postura tem sido a habitual, expressando um naturalismo em filosofia, tese segundo a qual os termos com significado podem ser expressos em linguagem científica e, em última instância, em linguagem física. Repercussões a esse respeito podem ser visualizadas na interseção entre neurociência e filosofia, por exemplo.

tradicionalmente avaliativos que são indesvencilháveis de um componente factual. Uma quantidade importante de termos avaliativos não pode ser fatorada em um componente puramente avaliativo e outro puramente descritivo, de modo que, ainda que adotássemos uma versão de fato que esteja dependente de nossas descrições, certas descrições implicarão termos avaliativos¹⁹.

Se a dicotomia entre juízos factuais e juízos valorativos encontrava alguma legitimidade na divisão estanque entre os enunciados analíticos e os enunciados sintéticos, os argumentos em favor daquela dicotomia malogram com a demonstração de que esta última divisão deveria ser abandonada. Como demonstraremos na próxima seção, não é possível que tenhamos uma noção de fato sem que haja valores, o que Putnam chama de “imbricamento” entre fatos e valores.

2. A Dicotomia entre Juízos de Fato e Juízos de Valor

A rejeição do possível caráter cognitivo dos termos avaliativos por parte do positivismo lógico ocorre porque a sua noção de fato, herdeira do empirismo humeano, não admite, pelos seus critérios de aceitabilidade racional, a intrusão de valores. Esses critérios de aceitabilidade racional são aqueles que dizem respeito às condições estabelecidas para aceitarmos ou não um enunciado. Ainda que admitamos que a alteração da noção de fato do positivismo lógico²⁰ permita a inclusão de enunciados que, tomados individualmente, não são factuais, isso não resolve o problema, uma vez que a distinção analítico/sintético mostra-se colapsada. Dessa forma, as sucessivas mudanças na noção de fato pelo Positivismo Lógico, bem como o esboroamento da dicotomia analítico/sintético, deixou um vácuo no tocante aos aspectos cognitivos tanto dos fatos, quanto dos valores.

Passaremos à apresentação da posição de Putnam acerca da distinção fato/valor, que fora inflada até tornar-se uma dicotomia. Ele apresenta duas linhas argumentativas. A primeira afirma que nossas práticas cognitivas instituídas, a saber, a prática de investigação científica, pressupõem valores. A segunda linha argumentativa procurará mostrar que enunciados com valores éticos não são destituídos de conteúdo factual, ou melhor, muitos enunciados éticos pressupõem considerações acerca de fatos, o que

¹⁹ A noção de que objetividade e fato devem ser igualados a descrição é rejeitada por Putnam, como veremos a seguir.

²⁰ *Op. cit.* 18

garante que não sejam meramente convencionais. Esses dois argumentos perfazem a noção de imbricamento entre enunciados de fato e enunciados de valor, que terá como consequência aquilo que Putnam tem chamado de colapso da dicotomia fato/valor, o que deixará claro que nossos ditos enunciados factuais são avaliativos.

A primeira linha argumentativa afirma que “os procedimentos através dos quais decidimos da aceitabilidade de uma teoria científica têm que ver com o fato de a teoria científica como um todo apresentar ou não certas ‘virtudes’” (PUTNAM, 1992, p. 173). A premissa do argumento estabelece que a “verdade”, para a ciência, é um conceito puramente formal, que por si só nada diz com respeito ao conteúdo dos enunciados. Putnam declara que se quisermos saber algo acerca do conceito de “verdade” que um determinado discurso adere precisaremos esclarecer, de antemão, quais são os padrões de aceitabilidade racional que estão pressupostos em tal discurso. Isso significa dizer, do ponto de vista da verdade, que o conteúdo de um enunciado está sujeito a determinantes, que são, em última análise, os critérios utilizados no tocante à aceitabilidade racional.

Desse modo, caso a ciência almeje asseverar um enunciado qualquer, tal como “o bóson de Higgs é uma partícula subatômica”, deverá fazê-lo somente se “o bóson de Higgs é uma partícula subatômica”. Se esse não for o caso, então o enunciado científico é descartado. De acordo com isso, Putnam afirma que os enunciados científicos são, na perspectiva da verdade, formais, pelo menos até esse ponto das nossas considerações, pois dizer unicamente que a ciência procura a verdade, sem apresentar as condições de aceitabilidade racional, é dizer enunciados puramente formais (1992, p. 169).

Para Putnam, as noções de verdade e aceitabilidade racional andam de mãos dadas (1992, p. 249). Com isso, a iniciativa da investigação científica é configurar uma imagem do mundo conforme certos critérios (rigorosos, metodologicamente construídos), em que a palavra “conforme” exige que esses critérios de aceitabilidade racional sejam satisfeitos na investigação própria à ciência. Putnam conclui, com relação a isso, que, caso consideremos que o objetivo da ciência é chegar a alguma espécie de verdade, “a própria verdade obtém a sua vida dos nossos critérios de aceitabilidade racional”. (1992, p. 170) E aqui chegamos a um ponto de inflexão no argumento de Putnam, segundo o qual tais critérios irão nos revelar os valores que estão implícitos na forma investigativa e metodológica da ciência.

Putnam nos afirma que há conceitos implícitos à investigação científica, tais como “coerência”, “abrangência”, “simplicidade”, “plausibilidade”, etc. Esses conceitos seriam centrais para a investigação científica, pois sem eles não poderíamos formar uma teoria qualquer, muito menos cotejar teorias para sabermos qual é a melhor, ou que teoria é mais apropriada para explicar determinado conjunto de eventos, de modo que funcionam como os denominados critérios de aceitabilidade. Em uma palavra, os conceitos trazidos aqui são condições para que o empreendimento científico possa ser consistente e significativo, e sem eles a tarefa científica estaria malograda de antemão. Putnam diz que esses conceitos são valores epistêmicos.

De acordo com essa afirmação, pretendemos nos aproximar da verdade do mundo fazendo uso das teorias que escolhemos, e essas teorias possuem os valores, já listados, de “coerência”, “preditibilidade”, por exemplo. O que Putnam alega é que esses valores são componentes através dos quais nós conseguimos obter nossas descrições do mundo. Podemos compreender essa alegação da seguinte maneira: nossas descrições do mundo estão comprometidas com os valores, donde concluiremos, por um lado, que sem os valores não conseguiríamos formular uma visão de mundo, de modo que precisamos deles para ser exitosos (mesmo) em nossas pretensões cognitivas mais sofisticadas, como é o caso das teorias científicas (PUTNAM, 2008a, p.52); e, por outro lado, concluiremos que uma justificação externa desses mesmos valores está fora da alçada de confirmação pela via tradicional, segundo a qual deveríamos poder verificar sua objetividade por meio de uma correspondência com uma descrição correta da realidade.

Essa última afirmação acerca da descrição correta da realidade nos remete à posição do realismo²¹ tradicional, assim denominada por Putnam, para o qual deve haver uma correspondência entre a realidade e as nossas tentativas de representar essa realidade tal como ela é. Desse modo, Putnam nos diz que:

²¹ Segundo Simon Blackburn, “Um realista sobre um assunto A pode sustentar (I) que os tipos de coisas descritas por A existem; (II) que sua existência é independente de nós, ou que não são objetos produzidos pela nossa mente, linguagem ou esquema conceitual; (III) que as afirmações que fazemos em A não são redutíveis a outros tipos de afirmações e que afinal revelar ser acerca de outra coisa que não A; (IV) que as afirmações que fazemos em A têm condições de verdade, constituindo descrições inequívocas de aspectos do mundo, sendo verdadeiras ou falsas em função dos fatos do mundo; (V) que conseguimos obter verdades acerca de A, e que é apropriado acreditar completamente no que afirmamos em A”. (BLACKBURN, 1997, p. 335)

...se o realismo metafísico (tradicional) estiver certo, e se se pudesse ver o objetivo da ciência simplesmente como tentar conseguir que o nosso mundo nocional “correspondesse” ao mundo em si, então podia alegar-se que estamos interessados na coerência, abrangência, simplicidade funcional e eficácia instrumental apenas porque estas são instrumentos com o fim de realiza esta “correspondência”. Mas a noção de correspondência transcendental e o mundo é em si um contrassenso. (PUTNAM, 1992, p. 174).

Putnam programa, a partir dessa consideração, encetar uma crítica ao realismo tradicional. Caso não o fizermos, teremos a permanência de problemas que gostaríamos de evitar, quais sejam, a noção de que os conceitos implícitos à investigação científica, os valores epistêmicos, são meramente convencionais, não sendo necessários senão para atender os nossos desideratos de correspondência à realidade unívoca; e a convicção de que “objetividade” é o mesmo que correspondência a objetos. Isso excluiria, de uma vez por todas, a tentativa de compreender os valores em seu aspecto cognitivo.

Destarte, para que possamos conferir aos enunciados avaliativos o caráter de cognitivos, devemos empreender uma reformulação à noção tradicional de objetividade, que é arrendatária dos pressupostos do realismo tradicional. Ao criticar o realismo tradicional, Putnam irá nos fornecer instrumentos para alterarmos a noção de objetividade. E, em não sendo objetivo somente o que é factual, como já tínhamos visto e criticado por outra via, depois das considerações sobre o realismo tradicional, a noção de fato surge como interdependente de nossa linguagem, que é, a uma só vez, descritiva e avaliativa. Começemos, então, a crítica de Putnam ao realismo tradicional.

A crítica de Putnam ao realismo tradicional inicia-se com a identificação de que uma das suas questões problemáticas consiste em pensar-se, segundo suas próprias palavras, “que existe uma totalidade de Formas, ou Universais ou ‘propriedades’, estabelecida de uma vez por todas, e de que todos os significados possíveis de uma palavra correspondem a uma dessas formas ou Universais ou propriedades” (2008c, p. 19). Putnam observa que o problema aqui patenteado diz respeito ao significado. Pensa-se, em geral, que o significado das palavras designa uma propriedade a qual estas palavras referem-se, de modo que uma palavra tem como função apreender o mundo tal como ele é, com suas propriedades intrínsecas.

Putnam supõe o exemplo das espécies naturais, salientando que todas as amostras de ouro puro, bem como os objetos forjados com ouro puro, possuem em comum a propriedade de constituírem-se de isótopos do elemento atômico de número 79. Entretanto, Putnam está convencido de que a palavra *ouro* não é sinônima de

“elemento atômico de número 79”. Isso ocorre porque o significado cotidiano da palavra ‘ouro’ não pode ser definido como uma propriedade ou conjuntos de propriedades. Wittgenstein²² corrobora a ideia de que fazemos uso de vasta gama de palavras com respeito às quais não existem propriedades comuns a todos os objetos a que estas palavras se aplicam com correção.

Desse modo, estabelece-se uma visão metafísica segundo a qual a realidade é apresentada como “um conjunto fixo de objetos ‘independentes da linguagem’ e uma ‘relação’ fixa entre os termos e suas extensões” (PUTNAM, 1999, p. 102). Putnam nos traz a reflexão de que os objetos não são independentes da linguagem, ao mesmo tempo em que afirma que a imagem de mundo “não é o produto da nossa vontade – nem das nossas disposições para falar de determinadas maneiras” (1999, p. 103). Com isso, Putnam está a dizer que a linguagem exerce a função de elaborar um vocabulário que possa estabelecer uma conexão com o mundo, vocabulário que pode ser *melhor* elaborado, com o fim de referir-se *melhor* do que antes aos objetos.

Podemos dizer que Putnam quer enfatizar a importância da linguagem em todo o trâmite da tarefa de representar o mundo. Na visão da metafísica tradicional, julga-se que a linguagem é impassível do ponto de vista da representação. Isso quer dizer que, frente ao mundo, a tarefa linguística consiste em estabelecer a função de designadora, sendo esta linguagem, porém, neutra. A neutralidade linguística dos termos que compõem a realidade implica uma imagem de mundo em que o instrumental fornecido pela linguagem serve meramente como expressão verbal de nossas impressões e representações.

Outra questão suscitada pela crítica ao mencionado tipo de realismo tradicional “é o pressuposto confortável de que há uma totalidade definida de objetos que pode ser classificada, bem como uma totalidade definida de todas as propriedades” (2008, p.19). A crítica de Putnam à definição peremptória do universo de objetos que podem ser referidos, e a coleção fixa de propriedades, pode levantar a falsa objeção segundo a qual o mundo é totalmente construído pelos indivíduos²³. Todavia, Putnam observa que

²² Ver a obra *Investigações Filosóficas*, de Wittgenstein.

²³ Para maiores elucidações a essas críticas feitas a Putnam, ver BOGHOSSIAN, Paul. *Medo do conhecimento: contra o relativismo e o construtivismo*, 2012, pp. 61-63. São Paulo: editora senac. Esta contestação de Boghossian toma como exemplo a noção de descrição da realidade constante na obra de Putnam *Realismo de Rosto Humano*, capítulo 6, intitulado *Verdade e Convenção*. Falar-se-á a respeito dessas críticas quando comentarmos o caráter não relativista do realismo interno.

nossas pretensões de conhecimento se dirigem a uma realidade em geral independente dos indivíduos, ou melhor, dos falantes.

Embora Putnam corrobore a ideia de que há uma realidade independente dos falantes, ele está seguro de que “a reflexão sobre a experiência humana sugere que nem a forma de todas as alegações de conhecimento nem os modos pelos quais elas são responsáveis perante a realidade são estabelecidos de antemão e de uma vez por todas” (2008, p. 20). Pode-se afirmar, portanto, que apesar de Putnam supor uma realidade em geral independente da mente, não há, com relação ao projeto humano acerca do conhecimento, uma independência concernente ao modo e à forma de nossas alegações cognitivas. Se isto não estiver correto, podemos estabelecer a noção rejeitada por Putnam, segundo a qual existe um número delimitado de propriedades que devemos simplesmente “descortinar”, “descobrir” e, feito isso, usarmos tais propriedade para denotar os objetos constantes na realidade, que por sua vez são também delimitados e fixos.

Mas a que tipo de objetos o realista tradicional se refere? Putnam afirma que “As formas tradicionais de realismo estão comprometidas com a alegação de que faz sentido falar de uma totalidade estabelecida de todos os ‘objetos’ a que nossas proposições podem referir-se” (2008, p.20). Embora isso seja dito, a pergunta continua em aberto. Putnam interroga se um evento histórico, como a Segunda Grande Guerra, é um objeto. Do mesmo modo, o céu é um objeto? Ou a imagem que percebemos no espelho é um objeto? Teríamos que atentar para os critérios estabelecidos pelo metafísico tradicional e, assim, identificarmos o que seria um típico objeto nessa ontologia.

Putnam nos diz que o problema sobre objetos está posto antes mesmos de supormos a existência de entidades como as que postulamos acima, dado o antigo critério de individualidade de um objeto, qual seja, o de que suas partes acompanham o movimento do objeto quando este se deslocar. Putnam obtempera que há objetos de existência inquestionável que infringem este critério e afirma: “tenho em casa uma lamparina que viola este critério; a sombra diminui sempre que a lamparina se desloca! Será que a lamparina neste caso não é um objeto?” (2008c, p. 21).

A resposta pode estar na noção de soma mereológica, na qual os objetos são, eles próprios, aditamentos de partes de outros objetos. Putnam supõe que a soma mereológica poderia solucionar o problema da lamparina, explicando o porquê de ela ser um objeto. Porém, ele destaca que as consequências podem ser fortes e os resultados

muito caros: “Se a ‘soma’ de *quaisquer* dois (ou mais) objetos for um objeto; se houver até um ‘objeto’ formado pela minha orelha esquerda e o nariz do leitor, então *existe de fato* um objeto equivalente à soma mereológica do corpo da lamparina e da sombra” (2008c, p. 21).

Este exemplo levantado por somas mereológicas suscita uma modificação das concepções tradicionais acerca dos objetos e, mais ainda, da alegação de conhecimento possível que deles temos, e de que suas propriedades possíveis são passíveis de mudanças centrais. Segundo Putnam:

O realista tradicional parte do princípio de que os nomes gerais correspondem, de maneira mais ou menos unívoca, a várias “propriedade” de “objetos”, em algum sentido de “propriedade” e em algum sentido de “objeto” estabelecidos de uma vez por todas, e de que as alegações de conhecimento são simplesmente alegações sobre a distribuição dessas “propriedades nesses “objetos” (PUTNAM, 2008, p.22)

Putnam considera que o realista tradicional está completamente certo em um ponto: ele tem a acertada convicção de que a realidade é independente, e que as descrições que dela fazemos e chamamos de ‘mundo’ devem manter uma espécie de responsabilidade cognitiva. Em que pese essa imagem, não devemos a ela conectar a noção de que existe uma descrição ou descrições que estão postas de uma vez por todas, todas elas possíveis, e com as quais devemos nos referir, necessariamente, a essa mesma realidade. A afirmação seguinte de Putnam é que com esta imagem das descrições que a metafísica tradicional preconiza perdemos o mais importante e verdadeiro aporte do pragmatismo: “o de que a ‘descrição’ nunca é uma mera cópia e de que estamos sempre criando novas maneiras de a linguagem poder ser responsável perante a realidade”²⁴ (2008c, p. 22).

Embora essa contribuição possa ser tomada de vários modos, tal como o próprio James²⁵ e Putnam o fizeram, esse diz que não estamos habilitados a concluir o mesmo que aquele, a saber, que nós criamos, em parte, o mundo. Ora, para Putnam a existência de uma realidade independente de nós é algo do qual não podemos duvidar a sério, e o simples fato de assumir isso já o coloca na categoria daqueles que aderem ao realismo. A característica agora enfatizada é a recusa da chamada ingenuidade do realismo

²⁴ Esse consideração é, segundo Putnam, inspirada em William James.

²⁵ Ver o texto *O Empirismo Radical*, de William James. Esse texto está relacionado na obra *Pragmatismo e textos selecionados*, arrolado na bibliografia desse trabalho.

tradicional, segundo a qual existe uma totalidade fixa de objetos e propriedades em cuja relação descobrimos as propriedades para denotar os objetos.

O que motiva esta dispensa da ingenuidade do realismo tradicional é a compreensão de que nossa linguagem e seu evidente uso estão colocados desde sempre, e que não devemos transigir com a ideia de que as propriedades, expressadas pelos conceitos, modos de referências, formas de descrições, apresentação de significados, etc., são imutáveis. Putnam esclarece sua posição da seguinte maneira:

A ideia de que nossas palavras e nossa vida são constringidas por uma realidade exterior a nós desempenha um importante papel em nossa vida e deve ser respeitada. A fonte de confusão reside no erro filosófico vulgar de se supor que o termo *realidade* tem de referir-se a uma única supercoisa, em vez de considerar as formas como renegociamos incessantemente – e somos *forçados* a renegociar – nossa noção de realidade à medida que nossa linguagem e nossa vida se desenvolvem. (PUTNAM, 2008, 23)

A esta crítica segue-se naturalmente o que Putnam nos trará, a saber, a constatação da ininteligibilidade da afirmação, feita pelo realista tradicional, de que há “uma nítida linha divisória entre as propriedades que ‘descobrimos’ no mundo e as que ‘projetamos’ no mundo” (2008, p. 35). Esta crítica é moderadamente explícita no que Habermas tentou elucidar quando conclama a metáfora de que há um ‘livro da natureza’, ensejada pelas suas considerações sobre o realismo pós-metafísico²⁶.

Isso significa que a realidade não possui termos próprios que podemos captar por meio de um processo de apreensão sócio-misterioso e, destarte, configurar uma noção de mundo que esteja a ela adequadamente acoplada em um sentido apenas em segunda instância linguístico. Seria só em segunda instância linguístico, pois o processo de captura da realidade “tal como ela é” não se daria à nossa compreensão por meio linguístico.

Este é o problema que está pressuposto na forma de acoplagem (o modo como a linguagem se conecta ao mundo) do realismo tradicional decorrente do que Putnam diz acima, qual seja, o de passarmos da apreensão das propriedades extralinguísticas da realidade para a conceituação semântica do mundo. O que pode ser evidenciado na passagem da percepção da realidade para a sua intelecção (linguística). A ininteligibilidade da “linha divisória”, indistinta, não nos habilita a afirmar que o mundo não possui propriedades extrínsecas à mente humana. Esta conclusão seria uma

²⁶ Antecipando o pensamento que Habermas cita de Putnam, “não existe tal coisa como a linguagem própria do mundo”. A citação completa está no capítulo 3 deste trabalho, na página 50.

extrapolação com a qual Putnam não se coaduna, pois sua consequência imediata é a recusa da existência de uma realidade independente de nossas descrições e conceituações.

O que a noção de ininteligibilidade da afirmação da linha divisória, com relação às propriedades “descobertas” e às propriedades “projetadas”, enseja concluir, é que subjaz à interpretação da classe dessas propriedades o uso linguístico que nós criamos para nos referir a ambas. Se não há uma linguagem factual da natureza e, portanto, ela não possui termos intrínsecos com os quais devemos a ela nos referir, torna-se azado sugerir uma linguagem capaz de fazê-lo, linguagem esta que é consignada pelo uso que dela fazemos²⁷.

Pode-se dizer, em uníssono com Putnam, que o que liga a crítica do realismo tradicional à reformulação do conceito de fato é a limitação que as tradicionais dicotomias “juízos de fato *versus* juízos de valor” e “verdades de fato *versus* verdades analíticas” promovem. É justamente a tomada de posição do realismo tradicional que irá contaminar o conceito de fato, fazendo deste um conceito que, em sua versão moderna, somente refere-se a objetos que podem ser percebidos pelos sentidos, deixando os valores relegados à subjetividade plena.

Segundo Putnam, estas dicotomias estanques “corromperam nosso pensamento, tanto no raciocínio ético como na descrição do mundo, principalmente, por impedir-nos de ver como a avaliação e a descrição estão interconectadas e são interdependentes” (2008a, p. 15). A limitação se sobrepõe ao conteúdo filosófico e torna-se consagrada. O esforço de Putnam em buscar um novo tipo de realismo que se preste a uma explicação heterodoxa de como nos relacionamos com o mundo nos é importante porque ela necessariamente reformula aspectos da metafísica tradicional, lançando-lhe as bases para uma contestação do conceito de objetividade que se consolidou na história da filosofia, desde a filosofia moderna²⁸.

²⁷ Não haver uma linguagem própria à natureza, com a qual poderíamos nos referir a ela perfeitamente, caso conseguíssemos fazer uso de tal linguagem, não nos coloca na difícil situação na qual não podemos nos referir com precisão à realidade: apenas teremos, para descrever e denotar as propriedades da natureza, que fazer uso de uma terminologia que não é intrínseca à realidade, mas a qual ela se refere e por ela é motivada efetivamente. Podemos afirmar que os termos que *usamos* para nos referir à natureza lhe são extrínsecos, embora se refiram a ela, sempre com maior ou menor grau de precisão, e por ela são motivados. Por isso, a nós, a nós, a aceção linguística, cabe “renegociar” nossos termos de referência (PUTNAM, 2008c, p. 23).

²⁸ A dificuldade que se tem ao tentar conectar as questões clássicas do realismo tradicional a uma análise de sua influência nas questões de valor (ético) está posta pela falta de acuidade que temos, pois não

Passemos à análise sobre o pressuposto último do realismo tradicional criticado por Putnam, a saber, que há “uma relação fixa de ‘correspondência’ em termos da qual se supõe que a verdade seja definida” (2008c, p.35). Putnam nos lembra, após mencionar o pressuposto do realismo, que sua rejeição está caracterizada pela ininteligibilidade dessa afirmação. Este pressuposto parece resumir, de algum modo, todos os pressupostos anteriormente mencionados, cujos apanágios são: a fixidez das formas predicativas da linguagem, e a conseqüente fixidez das interações entre os aspectos da realidade e a mente; e a possibilidade de nos referirmos aos objetos do mundo não só de modo adequado, mas de modo unívoco, a que se segue uma capacidade de podermos fazer corresponder nossos enunciados linguísticos a tais objetos, desembocando em uma teoria da verdade por correspondência unívoca.

A rejeição, ou declaração de ininteligibilidade dos pressupostos da metafísica do realismo tradicional é, necessariamente, uma recusa da noção de verdade que destes pressupostos decorrem. Se considerarmos que a noção de fato tem estado intimamente relacionada com ao menos um desses pressupostos, estaremos afirmando que o ‘fato’ conecta-se àquilo que no realismo tradicional chama-se objeto. Apesar de toda a discussão em torno do que é um objeto, sabemos que na história da filosofia há uma relação de proximidade indesejável entre fato e objeto, pois só poderia ser factual o que é objetual. Ao contrário, o subjetivo não poderia ser um fato. A crítica de Putnam é dirigida ao realismo tradicional e à absorção de algumas das decorrências desse realismo pelo empirismo humeano – como a noção tradicional de fato²⁹ e a representação dos objetos como cópia, proporcionada a partir do aparato sensorial.

Como vimos, para localizar o lugar dos valores e dos fatos na dicotomia tivemos que reorientar essas noções. Putnam nos apresenta uma crítica ao realismo tradicional, mas não nega que possamos almejar um tipo de adequação ao mundo empírico:

damos importância suficiente àquilo que as une – trata-se de uma polarização, também estanque, entre objetividade e subjetividade. Uma tomada de posição inovadora com relação ao realismo implicará outro ponto de vista concernente à objetividade, pois teremos que atenuar, no caso específico de Putnam, a linha entre a objetividade e a subjetividade para darmos conta dos aspectos perceptivos com relação ao mundo ou, em outras palavras, para conseguirmos explicar a relação entre a mente e o mundo. Está patente que nos interessa enfatizar os tópicos que poderão nos fornecer o instrumental necessário para seguir o argumento de Putnam, e que a linguagem exerce um papel fundamental que estará no centro da compreensão putnamiana da verdade. Temos, neste momento, a sua crítica ao realismo tradicional.

²⁹ Corroborada, posteriormente, por características típicas do empirismo, como, por exemplo, os dados dos sentidos.

... usamos os nossos critérios de aceitabilidade racional para construir um quadro teórico do “mundo empírico”, e então à medida que essa imagem se desenvolve revemos os nossos próprios critérios de aceitabilidade racional à luz desse quadro e assim por diante, e assim por diante sempre (PUTNAM, 1992, p. 174).

Podemos perceber que um dos resultados da crítica de Putnam ao realismo tradicional é que nossa linguagem pode ser indefinidamente melhorada para nos referirmos ao mundo empírico, mundo esse externo e independente de nós. Embora essa independência (enquanto existência do mundo externa à mente humana) seja algo patente, a respeito do qual Putnam não contrapõe nenhum argumento, ele sublinha o papel da linguagem no mundo real. Desse modo, o mundo real, formado pelos objetos aos quais nos referimos, é moldado segundo nossos critérios de aceitabilidade racional, propostos por nossa linguagem.

A conclusão a que Putnam chega é a seguinte: nosso vocabulário é como que constrangido pelo mundo empírico, e não é uma criação aleatória e unilateral, pois o mundo participa da formação desse vocabulário. Porém, nosso enquadramento do mundo real não é indiferente e impassível a esse mesmo vocabulário e, uma vez que os nossos valores são a lente através das quais nós podemos visualizá-lo, esse enquadramento do mundo real depende dos nossos valores. Em suma, o mundo externo à nossa mente não depende de nossas considerações concernentes a ele para existir, mas o mundo, com seus objetos, depende de nossos valores para ser formado, haja vista que não existe uma linguagem independente de nós no mundo.

Ademais, Putnam declara que ao menos alguns dos valores epistêmicos têm que ser objetivos. Ele argumenta que aceitar racionalmente um determinado enunciado depende do pano de fundo constituído por valores epistêmicos, pois os enunciados cognitivos não podem ser formulados, significativamente, sem a adesão a termos valorativos (PUTNAM, 1992, p. 175). O imbricamento entre as noções de fato e valor nega que os enunciados factuais possam ser neutros. Afinal, a tese do imbricamento entre as noções de fato e valor repousa na revisão dos pressupostos do realismo tradicional, de modo que seria despropositado afirmar uma linguagem do mundo inteiramente factual, independente de nossos usos e aplicações conceituais (PUTNAM, 2008a, p. 147).

Se aceitarmos a crítica ao realismo tradicional em pelo menos uma de suas considerações, o teremos suficientemente fragilizado para podermos afirmar, depois de tudo exposto, que o modo de falarmos sobre a realidade exterior depende de conceitos

por nós aplicados a ela. Não obstante, esses conceitos são suscitados pela pressão dessa realidade exterior que, assim, corrobora sua formação, e, portanto, não são meramente casuais ou convencionais. O fato de nossos usos e aplicações de conceitos se referirem a uma realidade exterior independente de nós quer dizer, a uma só vez, que não há algo como propriedades de objetos independentes de nossa linguagem (2008a, p. 23); e também que nossa linguagem conceitual refere-se a características dessa realidade exterior (PUTNAM, 1992, p. 176).

De acordo com isso, podemos dizer que termos avaliativos como “coerente”, “simples”, etc., não são destituídos de uma aplicação que se relaciona com a realidade, visto que no tocante à aplicação desses termos há uma importante “adequação” a essa realidade mesma. Aplicar um conceito é, dessa maneira, adequá-lo àquilo que é referido não unicamente de modo individual (o que culminaria em um subjetivismo de tipo solipsista), mas também de modo reconhecível, o que significa que a aplicação deve poder ser compreendida por outros usuários da linguagem, e não tão-somente pelo indivíduo que realiza (ou aciona) a aplicação imediata (1992, pp. 158-159).

Portanto, a tarefa de aplicar um termo avaliativo à realidade não é uma tarefa apenas subjetiva, pois, caso o fosse, poderíamos dizer que uma teoria qualquer é “coerente, para o indivíduo x ” e “não é coerente, para o indivíduo y ”. Porém, isso contraria a própria natureza do termo “coerente”, pois há nesse termo uma reivindicação que não é meramente a possibilidade de uso privado, mas, antes, a reivindicação de uma aplicação conceitual que possa ser reconhecida por indivíduos em seu caráter intersubjetivo.

3. Conceitos Éticos Espessos: a conclusão

Concluiremos, assim, que crítica ao realismo tradicional nos conduz à percepção do quão importante e relevante são para a constituição da noção de fato os critérios de aceitabilidade racional. Os valores são, eles mesmos, partes necessárias sem as quais não teríamos um mundo de fatos (1992, p. 249). E esses valores são o centro da noção da aceitabilidade racional, justamente o que nos habilita dizer de um enunciado se ele é verdadeiro ou falso.

Putnam diz que um problema decorrente da dicotomia entre enunciados factuais e avaliativos é considerar separadas, de forma estanque, a nossa linguagem descritiva de nossa linguagem usual, que objetiva elogiar ou censurar (PUTNAM, 1992, 178). Por isso, habitualmente se tem suposto que há um tipo de linguagem cotidiana, que possui um vocabulário descritivo, que se refere somente a um mundo formado por fatos, e um vocabulário com termos propriamente avaliativos, do qual nos utilizamos quando queremos prescrever ou incentivar um comportamento, conquanto seja não cognitivo. Ambos os tipos de vocabulários seriam, nessa configuração, independentes entre si.

Desse modo, os conceitos éticos fariam parte do vocabulário avaliativo, pois tradicionalmente tomamos conceitos éticos como destituídos de racionalidade (em seu aspecto cognitivo, algo que subjaz à dicotomia fato/valor). Entretanto, a tarefa a que Putnam se propõe tenciona revelar o caráter do imbricamento do fato e do valor nas questões éticas, expondo, assim, a generalidade do colapso da dicotomia ora tratada. Os conceitos éticos abstratos, também chamados de tênues, são aqueles que utilizamos para dizer de uma ação se ela é correta ou boa. No caso, “correto” e “bom” são conceitos éticos abstratos (PUTNAM, 1999a, p. 240).

A característica de tais conceitos é não possuírem um componente descritivo independente das ações. Assim, o emprego dos termos éticos abstratos está relacionado aos enunciados como valores éticos, como, por exemplo, na frase “‘Não seja indelicado’ é uma prescrição boa”. Nesse exemplo, o termo ético abstrato não possui um componente descritivo independente, antes, depende do que o termo “indelicado” quer significar. Só a partir de uma consideração ao componente descritivo intrínseco ao termo “indelicado” é que o conceito ético abstrato pode ser aplicado, o que revela que sua aplicação é feita em segunda instância, pois depende do conteúdo do termo em questão.

Segundo Putnam, é justamente a classe de termos éticos exemplificada por “indelicado” que torna evidente o imbricamento entre juízos de fato e juízos de valor. Isso ocorre porque não podemos “fatorar”, ou seja, distinguir de modo estanque os componentes descritivo e avaliativo dessa classe de termos éticos (PUTNAM, 1992, p. 252). Essa classe de termos éticos foi denominada conceitos éticos espessos (ou densos). A nomenclatura se deve ao fato de que em certos casos os valores *são*

descrições e, portanto, ao invés de valores e descrições contrastarem, eles se sobrepõem e coincidem (PUTNAM, 2004, p. 74).

A forma mais direta de explicar os conceitos éticos espessos é com a colocação de um exemplo em que é patenteado que os dois componentes estão imbricados. Segundo esse exemplo, quando se diz que determinado indivíduo é cruel temos que (1) há uma importante tentativa de caracterização de um comportamento e (2) tal comportamento está, quando assim caracterizado, expondo traços do indivíduo que nos fornecem elementos avaliativos, que não podem ser excetuados do conteúdo descritivo do termo.

Desse modo, seria curioso afirmar que um indivíduo é cruel e bondoso, ao mesmo tempo, pois as características descritivas para um indivíduo cruel presumem considerações avaliativas, de tal modo inerentes, que seria impossível acrescentar-lhe a virtude da bondade. Igualmente, pode-se afirmar que os componentes descritivos e avaliativos, no tocante aos conceitos éticos espessos, são indissociáveis, por força de a descrição ser interdependente do componente avaliativo. Em sendo a descrição interdependente do componente avaliativo, e ante todo o exposto no capítulo, pode-se concluir que a tentativa de consolidar a dicotomia fato/valor utilizando a estratégia da tese dos dois componentes é malograda.

Segundo essa tese, não há uma interdependência dos componentes descritivo e avaliativo, e cada um deles está desconectado do outro. Essa perspectiva malogra porque seu sucesso depende de uma noção de fato que foi amplamente criticada nesse trabalho, qual seja, a noção segundo a qual somente fatos podem ser objetivamente predicados, produzindo o resultado de que a descrição é uma característica de objetos aos quais os predicados podem ser referidos. Quando superamos o realismo tradicional, a noção de objetividade é reformulada, uma vez que os objetos não estão acabados de modo peremptório, o que gera a consequência de termos de participar do processo de formação daquilo que prestará aos enunciados seu caráter cognitivo.

O imbricamento, como se pode perceber, irá relacionar de modo próximo e interdependente os componentes factuais e avaliativos. É esse expediente que irá prestar aos enunciados éticos a cognitividade aqui almejada, haja vista a mudança na noção de fato ser uma consequência das críticas lançadas. Para o estabelecimento do fato a descrição é insuficiente e precária. O componente avaliativo revelou-se, como demonstramos, precípuo para que o fato possa ser considerado. Temos, desse modo, a

objetividade por meio da cognitividade alcançada, e podemos, então falar quanto a um enunciado ético se ele é verdadeiro ou falso.

O MEDO E A ORIGEM DO ESTADO EM THOMAS HOBBS

Gerson Vasconcelos Luz¹

RESUMO: Thomas Hobbes apresenta o medo como uma paixão fundamental para a origem do Estado. Por tendência natural, o homem prefere a liberdade do estado de natureza à vida em sociedade. Aceita a restrição por dois motivos centrais, medo e auto conservação. Não é qualquer tipologia de temor, mas o medo recíproco da morte violenta. Com contrato que estabelece o Estado o medo recíproco é neutralizado e os homens convivem em paz. Porém, a vida em sociedade requer outro tipo de medo, o temor à espada da justiça.

PALAVRAS-CHAVE: medo; razão; origem; Estado.

Abstract: Thomas Hobbes has fear as a fundamental passion for the State of origin. For natural tendency, the man prefers the freedom of the state of nature to life in society. Accepts the restriction for two central reasons, fear and self-preservation. It is not any fear of typology, but mutual fear of violent death. With contract establishing the State reciprocal fear is neutralized and the men live together in peace. However, life in society requires another type of fear, fear of the sword of justice.

Keywords: fear; reason; origin; State.

Segundo Thomas Hobbes (2003, p. 143), o homem ama a liberdade em sentido natural. Em outras palavras, por natureza os indivíduos almejam a fazer tudo àquilo que bem entender em vista da autoconservação e da realização de seus interesses pessoais. A admissão do poder estatal como uma restrição a tais tendências só é aceita porque os indivíduos têm uma paixão em comum, a saber, o medo² recíproco da morte violenta e o anseio por uma vida mais confortável em relação às condições naturais. Assim, cabe investigar a importância do *argumento do medo* e sua importância no que tange a origem do Estado. No que diz respeito a isso, leiamos a seguinte passagem:

[...] Um homem pode por medo sujeitar-se a quem o ataca, ou pode atacá-lo; ou ainda, os homens podem se juntar para se sujeitar àquele sobre quem estão de acordo, por medo dos outros. Quando muitos homens se sujeitam conforme o primeiro modo, surge daí, como que naturalmente, um corpo político do qual procede a dominação paternal e despótica; e quando se sujeitam conforme o outro modo, por meio da mútua concordância entre muitos, o corpo político que formam é na maioria das vezes chamado de república, para distingui-lo do modo anterior, ainda que esse seja o nome geral dado a ambos (HOBBS, 2010, p. 101).

¹ Mestre em Filosofia pela UNIOESTE - Universidade Estadual do Oeste do Paraná.

² A causa do medo mútuo, na visão do autor, deriva-se da igualdade natural e da vontade recíproca de provocar danos (HOBBS, 1992, p. 33). Tuck situa a vontade mútua de causar dano na impossibilidade recíproca de se conhecer as intenções alheias. Observemos a sua argumentativa: “suponha que eu o veja caminhando pacificamente pela savana primitiva, assoviando e portando o seu tacape: você representa um perigo para mim: você pode muito bem pensar que não: sua disposição é inteiramente pacífica. Mas eu posso pensar que sim, e o exercício do meu direito à autopreservação depende apenas de minha avaliação da situação” (TUCK, 2001, p. 79).

Conforme a passagem supracitada há pelo menos dois modos de se discutir a questão da submissão. Primeiro caso: a submissão é aceita em razão do medo que os submetidos sentem em relação àquele que se acha em posição de domínio. Nesse caso, nota-se um problema: sendo a diferença de forças corporais e intelectivas pouco significativas para sustentar uma possível diferença entre indivíduos, como derivar do medo, a origem do poder estatal? Pela lógica dos argumentos, por medo da morte um homem se submete àquele a quem teme ou deve sua vida. Segundo caso: o poder comum deriva da união dos indivíduos por medo recíproco que se submetem à proteção de uma terceira pessoa (o soberano). Esse argumento é mais interessante, pois, considerando que os homens são naturalmente iguais, somente poderiam aceitar como soberano sobre si uma pessoa artificial.

O Estado se origina, em parte, da razão e, em parte, do medo. A razão constrói a pessoa artificial do soberano, e o medo da morte violenta faz com que os indivíduos submetam-se mediante acordo recíproco a tal artifício.

Como Hobbes deixa claro em *Elementos da Lei*, ambos os modos de estabelecimento do poder estatal podem ser denominados *república*. No primeiro caso tem-se o poder despótico ou dominação paternal e, no segundo, o domínio institucional. Em *Do Cidadão* (1992, p. 111), nosso autor afirma que a república por aquisição é *natural* enquanto a república por instituição é *política*.

O ponto fundamental para o estabelecimento do poder estatal é o consentimento de todos ou da maioria dos indivíduos. Isso está presente, seja na república por aquisição, seja na república por instituição, porque, em ambas as situações, os elementos fundadores da sociedade são os mesmos: a procura pela neutralização do medo da morte violenta e o desejo de vida confortável. Em ambas as possibilidades, uma vez instauradas o poder estatal, o soberano torna-se provedor da segurança e proteção em relação aos desideratos humanos contidos como pressupostos no contrato de autorização. Entretanto, a tese da sociedade natural, qual seja, da república por aquisição, parece ficar em segundo plano no sistema de hobbesiano. Isso já se faz notar no começo da primeira parte em *Do Cidadão*, onde o autor considera que aceitar a ideia de que o homem é um animal político por natureza equivale a um erro procedente de se investigar a natureza humana com muita superficialidade (HOBBS, 1992, p. 28).

O objeto de interesse, ao aceitar o poder estatal como restrição à liberdade a todas as coisas, não é outro senão a precaução em relação à proteção de interesses particulares de cada indivíduo. Trata-se, desse modo, de uma forma egoísta de procurar

a realização de desejos. Atenemos para as paixões que implicam condições de possibilidade para que os homens possam se reunir em forma de sociedade. Escreve Hobbes:

As paixões que fazem os homens tenderem para a paz são o medo da morte, o desejo daquelas coisas que são necessárias para a vida confortável e a esperança de as conseguir por meio do trabalho. E a razão sugere adequadas normas de paz, em torno das quais os homens podem chegar a um acordo (HOBBS, 2013, p. 111).

Nota-se que o abandono à vida em liberdade em sentido natural não se dá gratuitamente. A ideia de ganho econômico está presente. Viver numa condição na qual o risco de perder a vida ou de sofrer danos é menor significa um ganho econômico, pois ter acesso efetivo aos meios que permitem uma vida mais confortável constitui-se em benefícios. A necessidade de autoconservação e as paixões egoístas exigem que o indivíduo trabalhe visando sempre maximizar ganhos e minimizar perdas. Nesse sentido, conforme mostra Polin (1992, p. 94), “[...] a esperança de viver bem graças ao seu trabalho substitui a esperança de conservar sua vida defendendo-se pela violência”.

É notável que paire entre os homens certa predisposição para causar danos recíprocos. A razão central pela qual se age assim está na supervalorização de si, no colocar-se a si mesmo no foco das ações, no individualismo que as condições naturais favorecer.

Deve-se ressaltar que as razões pelas quais os homens aceitam viver sob poder comum não é a conservação e defesa da vida do outro ou o desejo de coisas que pode lhes trazer comodidade. O outro indivíduo está sempre em segundo plano, mas, antes de se concluir precipitadamente sobre as possíveis implicações morais da questão, devemos atentar para o fato de que essa espécie de zelo incondicional sobre si mesmo deriva da necessidade natural que se tem em relação à autoconservação, à necessidade de proteção à vida em estado de natureza.

Diante disso, indaga-se: uma vez estabelecido o poder soberano, por que os homens permanecem egoístas? Por que mantêm o individualismo? A resposta parece estar contida na ideia de a sociedade não visa modificar a natureza humana; visa permitir que cada qual realize seus objetivos pessoais sem entrar em confronto com seus concidadãos.

A lógica do comportamento humano é deduzida das razões pelas quais os homens se reúnem. Uma das melhores passagens para notarmos a argumentativa de

Hobbes em torno dessa questão se encontra em *Do Cidadão*, capítulo I. Leiamos alguns trechos:

Como, e com que desígnio, os homens se congregam, melhor se saberá observando-se aquelas coisas que fazem quando estão reunidos. Pois, quando se reúnem para comerciar, é evidente que cada um não o faz por consideração a seu próximo, porém apenas a seu negócio; se é para desempenhar algum ofício, uma certa amizade comercial se constitui, que tem em si mais de zelo (*jealousy*) que de verdadeiro amor, e por isso dela podem brotar facções, às vezes, mas boa vontade nunca; se for por prazer e recreação da mente, cada homem está afeito a se divertir mais com aquelas coisas que incitam à risada, razão porque pode [...] mais subir em sua própria opinião quando se compara com os defeitos e deficiências de outrem [...] (HOBBS, 1992, p. 29-30).

Nota-se que o móbil que faz com que cada qual se posicione em relação a fatores externos é o interesse próprio – aquilo que agrada em primeiro lugar a si mesmo. O outro indivíduo, assim como todo objeto de interesse, é visto como um meio para a satisfação da vontade particular. Prossegue o autor:

Assim constatamos que, o mais das vezes, nesse tipo de reunião ferimos os ausentes; sua vida inteira, todos os seus direitos e ações são examinados, julgados, condenados; é até mesmo muito raro que algum presente não receba alguma seta antes de partir, de modo que não é má a razão daquele que procura ser sempre o último a ir embora (HOBBS, 1992, p. 30).

Aparentemente, se o interesse particular não estiver em questão não há motivos para estabelecer sociedade alguma. Assim, as paixões permitem que os indivíduos se posicionem favoravelmente à sociedade ou – dependendo do interesse em jogo – que se movam em aversão. Isso se dá desse modo, porque, afirma Hobbes,

Embora os benefícios desta vida possam ser ampliados, e muito, graças à colaboração recíproca, contudo – como podem ser obtidos com mais facilidade pelo domínio, do que pela associação com outrem –, espero que ninguém vá duvidar de que, se fosse removido todo o medo, a natureza humana tenderia com muito mais avidez à dominação do que a construir uma sociedade (HOBBS, 1992, p. 32).

Deve-se ressaltar, a origem do Estado está no medo recíproco: “somente o medo da morte convence o homem a renunciar às honras e às vantagens do estado natural e a aceitar o Estado civil” (BOBBIO, 1991, p. 85). Entretanto, o medo, por si só, é insuficiente para fazer com que os homens instaurem o poder comum. Nesse sentido, é necessário que se recorra à reta razão, que, por meio das leis de natureza, mostra os meios apropriados à união. Entretanto, a razão por si só também não basta.

Angoulvent (1996, p.112-113) situa a origem do Estado noutro elemento. Para a comentadora, não é o medo de morrer que faz o homem procurar a sociedade. O

homem, segundo a intérprete, não é fraco. A questão é que os indivíduos temem a possibilidade da descontinuação, do desaparecimento da espécie. Entretanto, essa leitura de Angoulvent, ao que parece, está em desacordo com a filosofia de Hobbes. O sentimento de medo da morte reflete a preocupação que cada indivíduo tem em relação à sua própria existência. A esse respeito o nosso autor é bastante claro em *Do Cidadão* ainda no capítulo I, *todas as grandes e duradouras sociedades tem sua origem no medo recíproco* (HOBBS, 1992, p. 32).

A questão da origem do Estado precisa ser vista a partir do medo da morte e da razão humana.

Quanto à razão é necessário que se observe as três primeiras leis de natureza descritas no *Leviatã*. A primeira consiste na seguinte regra prudencial: “[...] *que todo homem deve se esforçar pela paz, na medida em que tenha esperança de a conseguir e, caso não a consiga, pode procurar e usar todas as ajudas e vantagens da guerra*” (HOBBS, 2003, p. 113). Se – na condição natural – os homens conseguissem viver em não haveria porque a paixão do medo se manifestar. Uma vez que, dado uma série de fatores ligados a questão do autointeresse, o estado de guerra é inevitável e o medo da morte violenta uma constante. Em suma, embora a razão indique a paz, no estado de natureza os indivíduos estão sempre recorrendo às *ajudas e vantagens da guerra*.

Da primeira lei deduz-se a segunda, cujo conteúdo orienta:

Que um homem concorde, quando outros também o façam, e na medida em que tal considere necessário para a paz e para a defesa de si mesmo, em resignar o seu direito a todas as coisas, contentando-se, em relação aos outros, com a mesma liberdade que aos outros homens permite em relação a si mesmo (HOBBS, 2013, p. 113).

Para que o direito a todas as coisas seja abandonado e a soberania autorizada pelos contratantes é preciso que se observe a terceira lei: “[...] *que os homens cumpram os pactos que celebrarem*” (HOBBS, 2003, p. 124). Não basta que os homens se esforçar e concordem em viverem em estado de paz, é necessário que cumpram os acordos celebrados.

Diante das contradições do estado de guerra, os homens optam pela instauração de um poder comum. O acordo é tácito. É como se cada uma das pessoas implicadas no acordo dissesse: “[...] *autorizo e transfiro o meu direito de me governar a mim mesmo a este homem, ou a esta assembleia de homens, com a condição de transferires para ele o teu direito, autorizando de uma maneira semelhante todas as suas ações*” (HOBBS,

2003, p. 147). Assim, os homens naturais instituem uma condição na qual aquele que recebe a autorização se torna soberano absoluto sobre as vontades dos contratantes.

O soberano, pessoa artificial, alma do corpo político, é instituído para oferecer segurança e proteção aos súditos. Assim, a lei suprema é a segurança e proteção do povo (HOBBS, 1992, p. 220). Observemos, na argumentativa a seguir, algumas das exigências básicas dos súditos:

1. Serem defendidos contra inimigos externos.
2. Ter preservada a paz em seu país.
3. Enriquecerem-se tanto quanto for compatível com a segurança pública.
4. Poderem desfrutar de uma liberdade inofensiva (HOBBS, 1992, 06, p. 223).

Com base nisso, a obrigação dos súditos em relação ao soberano se sustenta enquanto houver segurança e proteção³ (HOBBS, 2003, p. 188). Esgotadas essas duas razões pelas quais o Estado é instituído, esgota-se também o compromisso dos submetidos, que podem – se for o caso – passar a servir a outro senhor.

O medo recíproco da morte violenta que origina o Estado é neutralizado e os homens convivem em paz. Porém, a vida em sociedade requer outro tipo de medo, o medo da espada da justiça do soberano. Se, na natureza os indivíduos provocam mutuamente o medo da morte, no Estado o medo é provocado pelo fio da espada do soberano. Entre temer um homem e temer muitos homens – embora este não seja o motivo psicológico para o abandono ao estado de natureza, – escolhe-se a primeira possibilidade. Em suma, o Estado se origina da razão, que prescreve leis que orientam os homens a viver em paz com seus iguais, e do medo recíproco que faz com que os homens se submetam a uma terceira vontade.

BIBLIOGRAFIA

HOBBS, Thomas. *Do Cidadão*. Tradução de Renato Janine Ribeiro. São Paulo: Martins Fontes, 1992.

_____. *Leviatã*. Tradução de João Paulo Monteiro; Maria Beatriz Nizza da Silva. São Paulo: Martins Fontes, 2003.

_____. *Os Elementos da Lei Natural e Política*. Tradução de Bruno Simões. São Paulo: WMF Martins Fontes, 2010.

³ Embora o Estado tenha como finalidade oferecer segurança e proteção, Hobbes admite que seja impossível que essas coisas se estendam de modo completo aos cidadãos, de modo que ninguém corra risco de morte violenta, danos físicos ou morais. Mesmo assim, escreve o autor, “[...] pode-se providenciar que não haja causa justa para o medo” (HOBBS, 1992, p. 117).

ANGOULVENT, Anne-Laure. *Hobbes e a Moral Política*. Tradução de Alice Maria Cantuso. Campinas: Papyrus, 1996.

BOBBIO, Norberto. *Thomas Hobbes*. Tradução de Carlos Nelson Coutinho. Rio de Janeiro: Campus, 1991.

POLIN, Raymond. O mecanismo social no Estado civil. In: QUIRINO, Célia Galvão; SADEK, Maria Teresa (Org.). *O Pensamento Político Clássico* (Maquiavel, Hobbes, Locke, Montesquieu, Rousseau). São Paulo: T. A. Queiroz, 1980. p. 93-106.

TUCK, Richard. *Hobbes*. Tradução de Adail Ubirajara e Maria Stela Gonçalves. São Paulo: Loyola, 2001.

DA PREDILEÇÃO À EXIGÊNCIA: A CONCEPÇÃO KIERKEGAARDIANA DO AMOR ENQUANTO DEVER

José C. Lopes Marques¹

José Roberto Gomes da Costa²

RESUMO: Tendo como base a primeira parte de *As obras do amor*, texto publicado por Kierkegaard em 1847, o presente artigo pretende analisar o conceito kierkegaardiano de amor a partir de sua relação com o dever e da compreensão de *próximo* enquanto *reduplicação*. O texto tenta situar ainda a crítica feita pelo pensador da existência às formas de amor baseadas na afeição e na predileção e a vinculação feita por Kierkegaard entre o *ágape* e a eternidade.

PALAVRAS-CHAVE: Amor; Dever; Reduplicação; Ética; Cristianismo.

Résumé: Sur la base de la première partie des *Ouvres d'amour*, texte publié par Kierkegaard en 1847, cet article vise à analyser le concept de Kierkegaard de l'amour de sa relation avec le devoir et la compréhension tout en suivant le redoublement. Le texte tente encore au penseur critique adoptée par l'existence de formes de l'amour sur la base de l'affection et de tendresse et d'attachement faite par Kierkegaard entre *agape* et l'éternité.

Mots-clés: Amour; Devoir; Redoublement; Éthique; Christianisme.

INTRODUÇÃO

Dentre as muitas acusações atribuídas a Søren Kierkegaard, o epíteto de solipsista é um dos mais divulgados. A ênfase dada pelo filósofo dinamarquês em seus escritos ao indivíduo e à existência na sua singularidade é muitas vezes interpretada como uma negação implacável da alteridade e da coletividade. Tal compreensão, não obstante, deixa de considerar o pensamento kierkegaardiano em sua inteireza. Ignora, por exemplo, a categoria do próximo desenvolvida pelo Dinamarquês em um de seus textos mais célebres: *As obras do amor*. Neste escrito, que inaugura aquilo que alguns autores designam como o “Segundo percurso kierkegaardiano”³, percebe-se um ataque

¹ O autor é formado em Teologia e Filosofia com Especialização em Ensino de Filosofia pela Universidade Federal do Ceará (Campus Cariri), atualmente cursando o Mestrado em Filosofia na UFC. Atua como professor efetivo de Filosofia na rede pública do estado do Ceará e como professor de Teologia e Filosofia na Faculdade Batista do Cariri (FBC). É autor dos livros *Diário de sonhos do Doutor satírico* e *Cultivando a reciprocidade*. Contato: markvani18@yahoo.com.br

² O autor é formado em Teologia pela Faculdade de Teologia Integrada (FATIN) e em Filosofia pela Faculdade Evangélica do Meio Norte (FAEME). É professor no Seminário Batista de Fortaleza (SBFor) e coautor do livro *Cultivando a Reciprocidade* (EBR). Contato: jrobg2001@yahoo.com.br

³ Costuma-se empregar essa designação para identificar as obras produzidas após 1846, ano em que Kierkegaard publicou o seu famoso *Pós-Escrito conclusivo e não-científico às Migalhas filosóficas*. Nas obras publicadas nesse período Kierkegaard, em geral, rejeita a escrita pseudonímica. Além disso, nestes escritos a ênfase religiosa é ainda mais evidente.

veemente desferido por Kierkegaard a toda forma de amor baseada na predileção, no individualismo e no egoísmo. A questão é que para este filósofo o homem só é capaz de perceber de modo correto o próximo depois de fazer-se verdadeiramente indivíduo, mas nisto não há solipsismo.

Obviamente, a refutação do Kierkegaard solipsista exige uma ampla discussão sobre o modo como este pensador define o amor, principalmente do amor em sua inusitada relação com o dever, não esquecendo, além disso, da crítica tecida pelo Dinamarquês ao amor natural (*eros*) da concepção platônica e à amizade (*filia*) esboçada por Aristóteles, da compreensão do *próximo* enquanto *reduplicação* e a relação entre o *ágape* e a eternidade. Esta será a proposta desse artigo.

1 – A CRÍTICA KIERKEGAARDIANA AOS AMORES PREDILETIVOS

Pois o Cristianismo entende melhor do que qualquer poeta o que seja o amor e o que seja amar; justamente ele também sabe aquilo que escapa aos poetas, que o amor que eles cantam, ocultamente, é o amor de si [...] O amor natural ainda não é o eterno, ele é a bela vertigem da infinitude [...] Tanto o objeto do amor natural quanto o da amizade têm por isso o nome do amor de predileção. (KIERKEGAARD, **As obras do amor**).

Em um texto de 1847 publicado por Kierkegaard com o título *As obras do amor: Algumas considerações cristãs em forma de discurso*, o filósofo da existência analisa o sentido do amor em sua relação com o Cristianismo. Para isso, ele toma como referência o mandamento do amor enunciado pelo próprio Cristo nos evangelhos (Mt. 22:39) e o famoso Hino do amor do apóstolo Paulo (I Co. 13). Nesta obra, o pensador dinamarquês preocupa-se em estabelecer uma distinção entre o amor natural e a amizade das concepções platônica e aristotélica e o amor no sentido estritamente cristão. As duas primeiras formas são marcadas pela *predileção*, pela demonstração de afeição e por certa medida de prazer sensível. Neste sentido, o ato de amar está relacionado a nutrir determinado desejo por aquilo que consideramos apazível e que, em geral, tomamos por digno das nossas afeições mais íntimas. Estas formas de amor, no entender de Kierkegaard, não possuem o caráter de exigência, pois são amparadas por uma inclinação natural.

Qualquer discussão acerca do amor deve levar em consideração o caráter polissêmico desse conceito. Os gregos, por exemplo, empregavam pelo menos quatro termos para exprimi-lo. O vocábulo *storgê* identificava aquilo que pode ser traduzido modernamente como afeição, comumente identificado como o afeto dos pais em relação aos filhos e dos filhos pelos pais (LEWIS, 2005). O *eros* apontava para o amor na sua dimensão sensual, erótica e apetitiva. Já o termo *filia* identificava o que se costuma traduzir como amizade, marcado, sobretudo, pela lealdade e virtude dos amigos. O *ágape*, palavra menos frequente entre os gregos, era empregada para falar de uma forma de amor baseada na abnegação e no sacrifício. Se considerarmos o dinamarquês, veremos que neste idioma há duas palavras que podem ser traduzidas como amor. *Elskov* tem uma dimensão erótica, natural e apaixonada. *Kjerlighed* possui uma conotação mais abrangente e mais elevada do amor. Álvaro Valls (2007), no entanto, na sua introdução à tradução de *As obras do amor* de Kierkegaard, acrescenta que esta distinção não é tão simples quanto aparenta. Isso porque, algumas vezes, a primeira forma de amor nem sempre possui uma conotação sensual. A título de exemplo, o termo aparece no conhecido relato bíblico do bom samaritano. De qualquer forma, para Kierkegaard, *Kjerlighed* é o termo que traduz o amor *ágape* no sentido propriamente cristão, que será o objeto de estudo do pensador dinamarquês.

É preciso ressaltar, no entanto que, embora a preocupação de Kierkegaard em seu escrito seja esclarecer o sentido do *ágape* ensinado nas páginas do Novo Testamento, ao mesmo tempo, ele tece uma severa crítica a duas concepções de amor amplamente conhecidas na tradição filosófica, a saber: o amor natural (*eros*) vinculado à filosofia platônica e a amizade (*filia*) da concepção aristotélica, ambos os amores baseados na predileção e na preferência. Por conseguinte, a discussão kierkegaardiana exige ainda que minimamente a retomada desses conceitos.

O amor ensinado pelo Cristianismo é, nas palavras de Kierkegaard, o amor que não foi cantado pelos poetas. Impossível ler esta declaração de e não se lembrar do *Banquete*, célebre diálogo platônico onde os convivas de Agatão, poeta trágico ateniense, são desafiados a proferir um discurso em louvor a *Eros*. A partir do discurso pronunciado por Sócrates, percebe-se que o amor está ligado à *predileção*. Fazendo referência a uma personagem fictícia chamada Diotima, o mestre de Platão define o *Eros* como sendo o desejo de possuir sempre o que é bom e belo (PLATÃO, 2001). Em

consequência desse caráter prediletivo o amor só pode ser pensado como um sentimento que se nutre em direção a um objeto desejável. De fato, a relação entre o ato de amar e o prazer sensível torna impossível que o *Eros* seja direcionado a um objeto desprezível. Para Sócrates, um objeto nos atrai em virtude de sua beleza e nos repele em função de sua feiura. Dito de outro modo, aquilo que nos parece belo, incita o nosso desejo e isto nos leva a amá-lo. Por outro lado, o feio parece despertar o nosso repúdio, impedindo-nos de amar tal objeto. Além disso, conforme observa Campos (2013, p. 87) em sua descrição do *eros* platônico, “o desejo ingente de desfrutar do objeto querido exige que o mesmo seja de beleza compatível, razão porque, sempre que percebe o fim iminente do amor, o amante é surpreendido pela melancolia”.

À luz da crítica kierkegaardiana, a sensação aprazível que caracteriza o amor natural ao modo platônico diverge radicalmente do caráter exigente do *ágape* cristão. O poeta, na compreensão do Dinamarquês é tal qual uma criança mimada, incapaz de aceitar a ordem solene do amor (KIERKEGAARD, 2007). Por isso, como em uma espécie de fuga, emprega seus esforços apenas para cantar o amor. Não há no *ágape* a ternura condescendente, aquele prazer vibrante que, como um turbilhão febril, agita o coração dos amantes, aquela reciprocidade feliz e aconchegante capaz de satisfazer a fome inquietante de amar e ser amado. Que poeta, questiona-se o pensador da existência, ousaria dedicar um único verso para louvar uma forma tão antinatural de amor?

A discussão kierkegaardiana acerca do amor, como já destacado, passa também por uma crítica ao conceito de *filia* esboçado por Aristóteles. De fato, o caráter prediletivo da amizade é percebido com muita clareza no livro VIII da *Ética a Nicômaco*, o qual o filósofo grego dedica inteiramente para falar acerca dessa forma de amor. Na compreensão aristotélica, a reciprocidade é uma das marcas fundamentais da amizade. Segundo suas próprias palavras, “para serem amigas, as pessoas devem conhecer uma a outra desejando-se reciprocamente o bem, por uma das razões mencionadas” (ARISTÓTELES, 2001, VIII 1156a). De fato, a amizade marcada pelo vínculo recíproco, continua Aristóteles, “é perfeita tanto no que se refere à duração quanto a todos os outros aspectos, e nela cada um recebe do outro, em todos os sentidos, o mesmo que dá” (ARISTÓTELES, 2001, VIII 1157a). Em outras palavras, o ideal é que haja uma igualdade entre os amigos. Se na relação predominar a desigualdade, os

laços da amizade tendem a se romper. Como se vê, a *filia* aristotélica expressa muito bem o caráter prediletivo contra o qual Kierkegaard direciona a sua crítica. A existência da reciprocidade garante que a relação seja regida pelo princípio do prazer, terreno propício para crescer a afeição entre os amigos. Neste caso o amor é vivenciado não como expressão de um dever, mas em virtude de uma alentadora compensação. Obviamente, o autor dos *Discursos edificantes* tem duras palavras contra essa forma de amor. Seguindo de perto a declaração paulina segundo a qual “o amor não busca os seus próprios interesses” (I Co. 13:5), ele qualifica a amizade, em sua reciprocidade, como uma forma sofisticada de egoísmo. Para Kierkegaard (2007), o verdadeiro amor deve eliminar a troca, deve suprimir o “meu” e o “teu”, fato que não corre em uma relação marcada por um vínculo recíproco. Por que há a retribuição, na amizade, o “meu” e o “teu” acabam sempre retornando para os seus possuidores. “Deve-se vigiar”, já falara o Dinamarquês nos *Discursos edificantes* de 1843, “para não ceder à inclinação de tomar de novo aquilo que doamos” (KIERKEGAARD, 2010, p. 157).

Apesar de sua crítica ferrenha ao caráter prediletivo do *eros* e da *filia*, Kierkegaard está disposto a concordar que, em determinados momentos, tais amores podem ser assumidos pelo *ágape*. Em outras palavras, os amores naturais podem ser úteis desde que tenham como fundamento o amor ao *próximo*. Esta aproximação é sinalizada pelo autor dinamarquês no trecho seguinte:

Ama a pessoa amada fielmente e com ternura, mas deixa o amor ao próximo ser aquilo que santifica o pacto com Deus da união de vocês; ama teu amigo sinceramente e com dedicação, mas deixa o amor ao próximo ser aquilo que lhes ensina na amizade de um para com o outro a familiaridade com Deus (KIERKEGAARD, 2007, p. 83).

2 – O PRESSUPOSTO, O PRÓXIMO E A QUESTÃO DA REDUPLICAÇÃO

Supérfluo é provar aos homens que têm rosto. Supérfluo também seria demonstrar-lhes possuírem amor-próprio. O amor-próprio é o instrumento da nossa conservação. Assemelha-se ao instrumento da perpetuação da espécie. Necessitamo-lo. É nos caro. Deleita-nos. (VOLTAIRE, **Dicionário filosófico**).

Pois quando é dito “tu deves amar o próximo como a ti mesmo”, aí está contido o que é pressuposto, ou seja, que todo ser humano ama a si mesmo (KIERKEGAARD, **As obras do amor**).

Como já expressara Voltaire (2002) o amor a nós mesmos é a base de todas as nossas ações. O imperativo do amor, conforme salienta Kierkegaard parte de um pressuposto, tal pressuposto, por sua vez, aponta para o tipo de amor que deve ser direcionado ao *próximo*. O Cristianismo pressupõe que todo homem ama a si mesmo e depois acrescenta que ele deve dedicar ao *próximo* o mesmo amor que, naturalmente, dedica a si mesmo. A letra do mandamento indica que, em sua essência, o amor em relação ao outro não representa a mera anulação do amor-próprio, afinal de contas, não é dito que se deve amar o *próximo* em vez de si mesmo, mas como a si mesmo. De alguma forma, o *ágape* não descarta por completo o amor-próprio. Deve-se primeiramente amar-se a si mesmo para depois amar o próximo. Não obstante, alerta Kierkegaard (2007), é preciso amar-se a si mesmo da maneira correta. De fato, aquele que não aprendeu a amar a si mesmo do modo adequado também está impossibilitado de amar o *próximo*. E há, segundo o pensador da subjetividade, inúmeras formas de amar-se a si mesmo de forma equivocada, seja o leviano que se dispersa nas loucuras do instante, seja o melancólico que deseja livrar-se da vida, ou o atormentado que acredita fazer um favor a Deus ao se torturar (KIERKEGAARD, 2007).

Quem é o meu *próximo*? A pergunta emblemática feita pelo fariseu a Jesus no Evangelho é retomada por Kierkegaard e tratada cuidadosamente em *As obras do amor*. Em sua análise sobre esta questão, o pensador dinamarquês afirma que o *próximo* não deve ser pensado em um sentido seletivo ou exclusivista, isso porque, como já foi destacado, o que deve mover uma pessoa em relação ao *próximo* não é a predileção ou a afeição, mas o dever e a exigência. Assim, para o autor de *Doença para morte*, o *próximo* possui um caráter bem mais abrangente do que o amigo ou o amado no sentido erótico. Em termos bastante enfáticos ele afirma que, “se há apenas dois homens, o segundo é o *próximo*; se há milhões, cada um deles é próximo” (KIERKEGAARD, 2007, 37. Grifo nosso). O *próximo* é, por conseguinte, todos os homens. O conceito não admite exceção, nele estão englobados todos os homens, mesmos os mais indesejáveis. Ao modo kierkegaardiano:

O amor cristão ensina a amar todos os homens, absolutamente todos. Com a mesma força incondicional que o amor natural insiste em que só exista um único amado, com a mesma força incondicional o amor cristão leva para a direção oposta. Se se quer, em relação ao amor cristão excetuar uma única pessoa que não se quer amar, então tal amor não é igualmente amor cristão (KIERKEGAARD, 2007, p. 69).

Vê-se que, em Kierkegaard, o conceito de *próximo* está intimamente ligado com a compreensão do amor enquanto dever. Conforme destaca Jonas Roos (2007), ao se reconhecer o dever, implicado no conceito de amor cristão, aprende-se a ver o próximo em toda e qualquer pessoa. Essa compreensão, no entanto, leva-nos ao encontro do outro no sentido de nos tornarmos o *próximo* daquele que está em necessidade, como na parábola do samaritano misericordioso contada por Cristo. Neste ponto, mais uma vez Kierkegaard ressalta que o amor ao *próximo* é uma exclusividade do Cristianismo. O paganismo só conhecia o amor apaixonado baseado na *predileção*, o amor ao *próximo* lhe era totalmente estranho. A razão desse desconhecimento é simples: não existe nada de admirável e louvável no *próximo*, nada que possa ser cantado ou exaltado. Ele é, em certo sentido, o desprezível, o indesejado, o imperfeito. O Cristianismo compreende perfeitamente esta verdade, por isso, afirma Kierkegaard (2007), não pede que se admire o próximo, ordena simplesmente que ele seja amado.

Na compreensão kierkegaardiana, o conceito de *próximo* também aparece vinculado à ideia de *reduplicação*. A *reduplicação* consiste em uma profunda identidade com o *próximo*. Por meio dela, no entender do filósofo dinamarquês, compartilhamos a igualdade dos homens diante de Deus (KIERKEGAARD, 2007). Dito de outro modo, a partir desse conceito o amor é libertado de seu caráter egoísta. Ele representa o perfeito reconhecimento da alteridade, de qualquer alteridade. Além disso, a *reduplicação* descarta o exclusivismo que por vezes se manifesta em ciúme, pois, como afirma o Dinamarquês, “deve-se amar a cada um em particular, mas a ninguém exclusivamente” (KIERKEGAARD, 2007, p. 88). A exclusividade é, por assim dizer, o disfarce que oculta a face de um amor egoísta. A posse exclusiva é apenas mais uma expressão da *predileção*.

É preciso evitar, não obstante, interpretar o conceito de *próximo* em Kierkegaard em seu sentido apenas geral, pensá-lo de um modo distante e impessoal. O *próximo* é todo mundo, mas é também cada indivíduo. A rigor, não se ama o *próximo* em sentido geral, ama-se cada indivíduo e nele se evidencia o mandamento divino. Muitas vezes o todo, pela sua abrangência, acaba se convertendo em ninguém. Este escapismo não encontra fundamento na concepção kierkegaardiana. A rigor, eu não preciso perder tempo tentando encontrar o *próximo*, ele é o primeiro indivíduo que eu encontrar agonizando na necessidade, semelhante ao moribundo da parábola.

3 – O ESCÂNDALO DO DEVER AMAR

Sê sincero, ou, para que isso não te perturbe, eu mesmo quero admitir que muitas, muitas vezes em minha vida isto provocou em mim assombro e surpresa, e que às vezes me pareceu que o amor tudo perdia com isso, enquanto que na verdade ele tem tudo a ganhar. Sê sincero, confessa que talvez seja esse o caso da maioria das pessoas, que, quando leem as descrições ardentes dos poetas a respeito do amor e da amizade, estas parecem então algo de muito superior a este pobre: “tu deves amar”. (KIERKEGAARD, **As obras do amor**).

O amor no sentido propriamente cristão substitui a primazia do prazer e da predileção pela primazia da exigência e do dever. A radicalidade do Cristianismo está em ordenar que se ame. “O amor já existia no paganismo”, afirma Kierkegaard (2007, p. 41), “mais isto de se dever amar constitui uma mudança da eternidade”. Justamente, por estar ligado à *predileção* e a afeição natural, no paganismo, o amor não precisa ser ordenado. A rigor, não é necessário ordenar que alguém ame seu cônjuge, ou a um amigo que dirija tal sentimento a seu consorte. Tais sentimentos parecem surgir espontaneamente. O *ágape*, contudo, conserva um tom paradoxal: ele possui uma estreita relação com o dever. Diferente da visão pagã, onde o amor é cantado em versos decassílabos e exaltado nos romances sentimentais, o amor cristão é marcado pelo *escândalo*⁴. Nas palavras do autor de *Temor e tremor*, “o *escândalo* vigia o acesso para o essencial do Cristianismo” (KIERKEGAARD, 2007, p. 80. Grifo nosso). O *escândalo*, termo também traduzido como choque ou ofensa⁵ marca a reação do paganismo em relação ao cristianismo ao deparar-se com o *paradoxo*⁶ que envolve esta doutrina. Todavia, a radicalidade do Cristianismo, segundo Kierkegaard, vai além. Ele afirma que se deve amar e amar não somente o amável, mais aquele que não é digno de ser amado como o próprio inimigo. Não se trata, é claro de uma relação erótica, pois, neste caso, não há beleza no objeto amado (VALLS, 2012, p. 80). E nisto está a excelência do *ágape* cristão. Em um trocadilho kierkegaardiano, “o amor ao *próximo*

⁴ Segundo France Farago, o escândalo é o sofrimento da razão crucificada nas tensões e contradições da realidade ontológica que a precede na ordem do ser. Em um texto de 1850, denominado *Exercício do Cristianismo* e publicado com o pseudônimo Anticlimacus, Kierkegaard esclarece que o Cristianismo é sempre marcado pela possibilidade do escândalo, pois coloca a fé paradoxal como elemento fundamental da relação entre o homem e Deus, negando com isso a possibilidade de uma comunicação direta. O acesso a Deus, portanto, não se dá pela via da razão, mas pela via da fé.

⁵ A expressão é empregada por Ricardo Quadros Gouvêia em sua obra *Paixão pelo paradoxo*. O evangelho é ofensivo, por isso, é rejeitado pelo paganismo.

⁶ Em obras como *Migalhas filosóficas* e no *Pós-escrito conclusivo não conclusivo às migalhas filosóficas* Kierkegaard desenvolve o conceito de paradoxo absoluto, fazendo referência à doutrina da encarnação de Cristo. Ou seja, não há como compreender racionalmente o modo como o eterno se delimita no tempo.

possui todas as perfeições da eternidade exatamente por não se prender às imperfeições do objeto amado” (KIERKEGAARD, 2007, p. 55). De fato, argumenta Jonas Roos (2007), poderia parecer que o amor-dever tiraria do amor a sua beleza, criatividade e liberdade em relação à pessoa amada, tornando-se uma obrigação fria e enfadonha. No entanto, bem entendida, a riqueza do amor especificamente cristão está precisamente neste tu deves. O paganismo, contudo, não considera que isto seja um sinal de perfeição, por isso, escandaliza-se com o mandamento do amor.

O poeta e o Cristianismo explicam exatamente o oposto: o poeta idolatra a inclinação e tem então, já que só pensa no amor natural, toda razão ao dizer que seria a maior tolice e o discurso mais absurdo ordenar amar; o Cristianismo, que pensa sempre no amor cristão, tem também, contudo, toda razão quando destrona essa inclinação e coloca no lugar dela aquele “deves” (KIERKEGAARD, 2007, p. 70).

A essa altura, uma retomada do *Banquete* de Platão seria bastante esclarecedora. Para a sua festa, o poeta Agatão convidara apenas aquelas pessoas que, no seu entender, eram dignas de estar ali, como seus amigos mais íntimos e pessoas destacadas por sua sabedoria. Dentre os que não tinham sido dignos de receber o convite de Agatão encontrava-se Aristodemo. Como desejava participar daquela festa, ele roga a Sócrates que diga a todos que o convidou. “Espero que me desculpes lá chegando. Se não o fizeres, não terei coragem de dizer que fui sem convite, mas declararei que tu me convidaste” (PLATÃO, 2001, p. 96). Reconhecendo que não tem a dignidade suficiente para participar daquela celebração, Aristodemo recorre àquele que poderia justificar a sua indignidade. Já que estamos falando em banquete⁷, Kierkegaard ilustra o caráter exigente do amor fazendo alusão à parábola contada por Cristo de um homem que dá um banquete e convida não os familiares e amigos importantes, mas os cegos, aleijados e desprezados da sociedade. Que diferença entre o *eros* platônico e o ágape cristão! Tal atitude requer o exercício de um amor não marcado pela *predileção* e pela afeição, pois o convite seria feito a pessoas bem pouco desejáveis. Neste contexto, entra em cena a dimensão exigente do amor. Tais pessoas não possuem nada digno de ser amado. Não há prazer em amar o desprezível, mas é dever do cristão amá-lo mesmo assim, diria o pensador dinamarquês. A verdade é que, conforme destaca C. S. Lewis (2005, p. 111) “existe em cada ser humano algo que não pode ser amado naturalmente. Pedir que isso

⁷ Dentro do cristianismo primitivo o banquete tem o seu lugar de destaque. Conforme o relato joanino, em um banquete Cristo realizou seu primeiro milagre – transformou água em vinho nas bodas de Caná da Galileia; em um banquete o Mestre se despediu de seus discípulos antes de sua paixão; após a ressurreição, depois de um banquete na praia do mar da Galileia, renovou o ânimo dos discípulos; os cultos primitivos eram sempre acompanhados de banquetes: as chamadas festas ágapes ou festas de amor.

seja amado seria o mesmo que pedir que alguém goste do sabor embolorado do pão ou do barulho emitido pela furadeira elétrica”. Com um objeto não amável só é possível relacionar-se amorosamente por meio de um amor gracioso, dadivoso e desprovido de qualquer atitude interesseira. Eis a exata dimensão do *ágape* na compreensão kierkegaardiana. Na análise do pensador dinamarquês percebe-se claramente a distinção entre natureza e graça referendado pela tradição cristã, expressado, por exemplo, por Tomás de Kempis quando afirma que “a natureza trabalha por seu interesse e só atenta no lucro que lhe pode vir; a graça não considera o que lhe é útil e cômodo, mas o que a muitos é proveitoso” (KEMPIS, 2008, p. 166).

Deve-se acrescentar, não obstante, que a questão não reside em apenas alimentar as pessoas pobres. Essa refeição precisa ter a dignidade de uma grande festa. “Aquele que alimenta os pobres, mas não sobrepuja seus sentidos a ponto de chamar esta refeição de um banquete, só vê no pobre e no pequeno um inferior; aquele que dá um banquete vê no pobre e no pequeno o próximo – por mais ridículo que isso possa parecer aos olhos do mundo” (KIERKEGAARD, 2007, p. 104). Deve-se evitar, por conseguinte aquilo de Kierkegaard denomina no primeiro capítulo de sua obra como praticar as obras do amor de modo desamoroso. Ver no miserável uma forma de manifestar a nossa grandeza, ainda que seja por meio do saciar a sua fome é um modo desamoroso de praticar as obras do amor. A propósito, nos *Discursos edificantes* de 1843 o filósofo dinamarquês denuncia a atitude daqueles que estão dispostos a ajudar o necessitado, mas, ao mesmo tempo, exigem deste seu respeito, admiração e sujeição (KIERKEGAARD, 2010). Neste ponto, Pascal poderia ser invocado para reforçar a tese kierkegaardiana. Nas palavras do pensador francês, “apiedar-se dos desgraçados não é contra a concupiscência. Ao contrário, bem satisfeitos ficamos com render esse preito de amizade e angariar uma reputação de ternura sem nada em troca” (PASCAL, 2005, p. 142). A essência graciosa do *ágape* consiste em uma espécie de rebaixamento daquele que ama e, ao mesmo tempo, em uma tentativa de anulação da miséria do objeto amado.

4 – O AMOR QUE SE SUBMETEU À TRANSFORMAÇÃO DA ETERNIDADE

Os amores naturais só podem ter esperança de eternidade na medida em que se permitirem ser assumidos pela eternidade da Caridade, e esse processo sempre envolve uma espécie de morte. (C. S. LEWIS, Os quatro amores).

Quando o amor submeteu-se a mudança da eternidade, em se tornando um dever, aí ele adquiriu continuidade, e daí seguiu-se de si mesmo que ele perdure. Pois não é evidente que aquilo que dura neste instante também venha a durar no próximo instante, mas é evidente que o contínuo perdura (KIERKEGAARD, As obras do amor).

Os amores espontâneos – *eros* e *filia* – por não terem se submetido à transformação da eternidade, estão sujeitos à alteração e podem converter-se no oposto do amor. Kierkegaard lista pelo menos três alterações às quais estão expostos tais amores. Em virtude de seu caráter fugaz, o amor pode converter-se em *ódio*. Nas palavras do Dinamarquês, “o *ódio* é um amor que se transformou no seu contrário, um amor que foi ao fundo. De alguma forma, o amor continua a queimar, mas a chama é a do *ódio*. Somente quando o amor tiver terminado de arder, a chama do *ódio* estará também extinta” (KIERKEGAARD, 2007, p. 51. Grifo nosso). Na ilustração kierkegaardiana, quando alguém afirma, por exemplo, “se não me amares então eu te odiarei”, expressa muito claramente um tipo de amor convertido em *ódio*. Tal alteração é própria das paixões fugazes que desaparecem com a mesma velocidade que se mostram. Esta mutação evidencia um amor não transformado pela eternidade, o único amor que seria capaz de dizer: “Mesmo se me odiares eu continuarei a te amar”.

Ademais, os amores imediatos estão sujeitos a se converterem em *ciúme* e, muito rapidamente, passa-se da mais envolvente felicidade ao mais pungente tormento. Falando acerca do *eros*, Lewis (2005) declara que ele pode trazer lágrimas aos olhos tanto quanto a dor. Para Kierkegaard, o ardor do *ciúme* é tão perigoso que muito facilmente pode tornar-se uma doença. O ciumento é aquele que perdeu a paz e a liberdade no ato de amar. Muito diferente do *ágape* que apazigua, que reconcilia, que edifica, o ciúme que brota a partir do amor imediato inquieta, separa e destrói. Dito de outro modo, o ciumento é torturado a cada instante pela ideia de ser ou não amado. Não há liberdade em seu amor, visto que ele se torna um prisioneiro da ansiedade e do medo. Quando o amor imediato converte-se em *ciúme*, surge a desconfiança. Por isso, para recorrer à analogia kierkegaardiana, “o ciumento vigia o objeto de seu amor com cem olhos, dominado pelo medo de perdê-lo. O *ágape*, não obstante, em sua simplicidade,

possui apenas um olho para o seu alvo” (KIERKEGAARD, 2007, p. 53). Isso porque, fazendo eco ao texto paulino, o amor tudo crê e, no entanto, jamais é iludido.

Os amores naturais podem, por fim, se transformar em *hábito*. Como nos lembra Kierkegaard, isso ocorre quando o amor perde seu fogo, sua alegria, seu prazer e originalidade, semelhante ao um rio que salta impetuoso do penhasco, mas vai enfraquecendo mais abaixo. A propósito o *hábito* é para o autor *Doença para morte* o mais pérfido dos inimigos dos amores imediatos. Diferente do *ódio* e do *ciúme*, ele jamais se permite mostrar como tal. Em uma analogia, “ele é semelhante ao animal sorrateiro que se alimenta do sangue de suas vítimas enquanto elas estão dormindo. Ele a refresca com o bater de suas asas tornando o seu sono ainda mais refrescante” (KIERKEGAARD, 2007, p. 55).

Segundo a concepção kierkegaardiana, o amor enquanto dever está intimamente vinculado à eternidade. Disso decorre que o problema da alteração ao qual estão expostos os amores espontâneos não lhe afetam. De fato, é só a partir de sua relação com a eternidade que o amor pode converter-se em dever e, por conta disso, está apto a permanecer. Como o pensador da existência afirma, “quando o amor submeteu-se à mudança da eternidade, em se tornando um dever, aí ele adquiriu continuidade e daí segue-se que ele perdure” (KIERKEGAARD, 2007, p. 49). Esta passagem está plenamente de acordo com aquilo que o filósofo de Copenhague já afirmara nos *Discursos edificantes* de 1843. Ali de modo assertivo, ele declarara que “toda dádiva boa e perfeita que um homem pode dar é amor, e, acerca dele, todos os homens em todas as épocas sabem que ele procede de cima” (KIERKEGAARD, 2010, p. 166). O exercício das obras do amor-dever, como já foi visto, é marcado por um caráter exigente. Por esta razão a vinculação entre o amor e a eternidade torna-se imperativa. Não sem razão, nos evangelhos, o mandamento do amor ao próximo é antecedido pelo mandamento do amor a Deus. Ou, como Kierkegaard (2010, p. 95) diria nos *Discursos edificantes*, “o segredo do amor terreal consiste em que leva sobre si o selo do amor de Deus”. Sem esta relação, o amor pode facilmente, converter-se em desespero. Todo aquele que tentar quebrar esse vínculo haverá de perder a essência do amor cristão. Como afirma Viallaneix (1977), o amor a Deus e o amor ao *próximo* são como duas portas que se abrem ao mesmo tempo, se uma estiver fechada, a outra também estará. Considerando a centralidade dessa relação, Kierkegaard desenvolve o conceito de

triplicidade. Para ele, quando o amor não está ligado à eternidade só é possível falar em reciprocidade, um relacionamento marcado pela troca entre o amante e o amado. Quando, porém, o amor passou pela transformação da eternidade não há mais apenas os dois, mas três: o amante, o amado e Deus, a própria essência e fundamento do amor. O amor que não está amparado pela *triplicidade*, mas baseia-se apenas na duplicidade da reciprocidade, mantém as portas abertas para o surgimento do amor egoístico já descartado pelo filósofo dinamarquês. Na superação do amor egoístico bem como na relação com a eternidade, é indispensável a *seriedade*. Conceito central do pensamento kierkegaardiano, a *seriedade* consiste em usar a vontade para dominar a si mesmo, tornando-se o que estava destinado a ser desde a eternidade e exprimir a eterna beatitude em cada ação de forma que o existente, existindo, transforme tudo na sua existência como prova de respeito ao Bem-Supremo (ALMEIDA, 2009, p. 50).

Somente quando o amor experimentou a mudança da eternidade tornando-se dever ele está eternamente assegurado. Esta segurança, nos diz o Dinamarquês, previne o *ágape* de ser afetado em três sentidos. Antes de tudo, o amor que submeteu-se à transformação da eternidade, está devidamente assegurado contra a *angústia*. Em um texto de 1844, sob o pseudônimo Vigilius Haufniensis, Kierkegaard já discutira amplamente acerca do conceito de *angústia*, enumerando as várias formas em que este estado se expressa e a sua relação com o *pecado* e a *liberdade*. Aqui o seu objetivo é bem mais modesto, restringindo-se apenas a demonstrar a vinculação que poderia existir entre a *angústia* e o amor, caso este não tenha se submetido à transformação da eternidade. Quando o amor não é dever, os amantes são facilmente assolados pela *angústia* ao se depararem com a possibilidade de sua alteração. A *angústia*, afirma Kierkegaard revela-se, sobretudo, no desejo que tanto o amante quanto o amigo têm de colocar seu amor à prova. Este anseio de que o sentimento seja provado revela a sua insegurança e, conseqüentemente, a *angústia* ao perceberem que seu amor não está completamente assegurado. A rigor, a prova não é, de jeito nenhum, a demonstração de segurança do amor. O provar, declara o pensador da subjetividade, “relaciona-se sempre com uma possibilidade, é de qualquer modo sempre possível que aquele que é provado não seja aprovado” (KIERKEGAARD, 2007, p. 50). Por outro lado, o *ágape* não precisa de nenhuma prova. Afinal de contas, que prova de permanência teria valor para o amor que traz em si o selo da eternidade? De fato, a conversão do *ágape* em dever

leva-o a prescindir de qualquer prova e essa é, no entanto, a suprema prova na qual ele é aprovado. Essa aprovação, por sua vez lança fora qualquer *angústia*.

O amor que experimentou a mudança da eternidade está assegurado não apenas contra a *angústia*, mas ele também está eternamente libertado em uma *feliz independência*. Em princípio soa desconcertante falar em *independência feliz* à luz de um amor que se define em sua relação com o dever. Como alguém que é obrigado a amar poderia ao mesmo tempo desfrutar dessa *independência feliz*? Antes de tudo, Kierkegaard tece uma crítica ácida ao pretensioso conceito de *independência* associado aos amores imediatos. *Independência* que não tem coragem para se comprometer é na verdade uma expressão de covardia. Além disso, continua o autor dinamarquês, ser independente não significa estar livre de necessidade. No amor proposto pelo autor de *O conceito de angústia*, a maior riqueza consiste exatamente em ter uma necessidade, sua maior liberdade consiste em sujeitar-se a uma lei. Ao modo kierkegaardiano:

O amor que se submeteu à transformação da eternidade em se tornando dever, e ama porque deve amar, é independente, tem a lei de sua existência na própria relação do amor para com o eterno. Este amor jamais pode tornar-se dependente no sentido não verdadeiro, pois a única coisa de que ele depende é o dever, e o dever é a única coisa que liberta. O amor imediato torna um ser humano livre, e no instante seguinte dependente (KIERKEGAARD, 2007, p. 56).

Como se vê, para Kierkegaard, somente no amor-dever encontra-se a feliz independência, a perfeita liberdade. Não obstante, o seguinte questionamento persiste: em que sentido esse amor-dever relaciona-se à *independência*? Um amor espontâneo, sem a prisão do dever, não parece muito mais apto a tornar-se independente? O filósofo de Copenhague responde a este questionamento mostrando que o dever liberta o amor porque o torna independente das contingências e vicissitudes do objeto amado. Assim é o *ágape*, ele não depende dessa ou daquela condição para dedicar-se ao objeto amado. Ele depende apenas do dever, a única coisa que o liberta, por isso, é eternamente independente. No comentário de Campos, o amor *crístico* não é definido por seu objeto, semelhante ao que ocorre em relação ao amor natural e à amizade. Ele é definido a partir de si mesmo e de sua relação com o eterno. Kierkegaard reitera o que já fora expresso por Agostinho. Em seu *Comentário a I Epístola de João* ele declara: “Ama e faz o que quiseres. Se te calas, cala-te movido pelo amor; se falas em tom alto, fala por amor; se corriges, corriges por amor; se perdoas por amor. Tem no fundo do coração a raiz do amor: dessa raiz não pode sair senão o bem” AGOSTINHO, 1989, VII, 8). Tal

não é o caso dos amores imediatos. Sua suposta *independência* é, na verdade, dependência, pois eles dependem dos ventos favoráveis da reciprocidade para poderem amar.

Por fim, na concepção kierkegaardiana, o amor que se submeteu à transformação da eternidade é capaz de fugir ao *desespero*. Em termos literais, somente “quando amar é dever, o amor está eterna e felizmente assegurado contra o *desespero*” (KIERKEGAARD, 2007, p. 58). Quando se ama em um sentido imediatista, pode-se facilmente ser atingido pela infelicidade e pelo *desespero*. Essas inquietações surgem tão logo se percebe o fim do amor. Como, no Cristianismo, o amor conserva uma estreita relação com a eternidade, não há razão para desesperar-se. Além disso, o *desespero* surge quando se relaciona com infinita paixão com o particular. O Cristianismo, por outro lado, defende que somente com o eterno deve-se relacionar com paixão infinita.

CONSIDERAÇÕES FINAIS

A partir das considerações acima, pode-se estabelecer dois pontos fundamentais. Em primeiro lugar, percebe-se que a acusação feita a Kierkegaard de solipsista é im procedente e implausível, pois não leva em consideração a categoria do *próximo* em *As obras do amor*. Assim, não seria forçoso afirmar que o pensador da subjetividade e da individualidade é também o pensador da alteridade. Visto que apenas quando o homem se faz indivíduo, e verdadeiro indivíduo singular é aquele que empregou a sua singularidade para dirigir-se ao Eterno, estará apto a dirigir-se ao outro de modo adequado. Dito de outro modo, não existe ninguém mais apto para reconhecer a alteridade amando o próximo do que aquele empregou sua singularidade para amar a Deus.

Em segundo lugar, não há contradição na relação entre amor e dever. Em outras palavras, o dever não é necessariamente uma negação da independência do amor. Como vimos, é possível pensar em uma independência mesmo a partir do comprometimento, do dever. O *ágape* é independente no sentido de ter autonomia para ser eternamente o que é. Ele é tão livre que as contingências do objeto amado não são capazes de alterá-lo e convertê-lo em seu oposto.

BIBLIOGRAFIA

AGOSTINHO. **Comentário da Primeira Epístola de São João**. São Paulo: Edições Paulinas, 1989.

ARISTÓTELES. **Ética a Nicômaco**. São Paulo: Martin Claret, 2001.

BÍBLIA. Português. **A Bíblia Sagrada**: Antigo e Novo Testamentos. Versão revista e atualizada de João Ferreira de Almeida. Barueri: Sociedade Bíblica do Brasil, 2000.

CAMPOS, Eduardo da Silveira. **O conceito de amor em Kierkegaard**: A concepção de Ágape como doação. 2013. 118 p. Dissertação (Mestrado em Filosofia). Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro- RJ.

FARAGO, France. **Compreender Kierkegaard**. Petrópolis: Vozes, 2005.

GOUVÊA, Ricardo Quadros. **A Palavra e o silêncio**: Kierkegaard e a relação dialética entre fé e razão em Temor e tremor. São Paulo: Alfarrabio: Custom, 2002.

_____. **Paixão pelo paradoxo**: Uma introdução a Kierkegaard. São Paulo: Fonte Editorial, 2006.

KIERKEGAARD, Søren. **As obras do amor**: Algumas considerações cristãs em forma de discursos. 2. ed. Petrópolis: Vozes, 2007.

_____. **Discursos edificantes, Três discursos para ocasiones supuestas**. Madrid: Trotta, 2010.

_____. **Ejecitación del cristianismo**. Madrid: Editorial Trotta, 2009.

_____. **O conceito de angústia**. 2. ed. Petrópolis: Vozes, 2010.

_____. **O desespero humano**. São Paulo: Abril Cultural, 1979.

KEMPIS, Tomás de. **Imitação de Cristo**. São Paulo: Hedra, 2008.

LEWIS, C, S. **Os quatro amores**. São Paulo: Martins fontes, 2005.

PASCAL, Blaise. **Pensamentos**. Rio de Janeiro: Ediouro, 2001.

PLATÃO. **Apologia de Sócrates e Banquete**. São Paulo: Martin Claret, 2001.

ROOS, Jonas. **Tornar-se cristão**: O paradoxo absoluto e a existência sob juízo e graça em Søren Kierkegaard. 2007. 247 p. Tese (Doutorado em Teologia). Escola Superior de Teologia (EST), São Leopoldo, RS.

VALLS, Álvaro. **Kierkegaard**: Cá entre nós. São Paulo: LiberArs, 2012.

VIALLANEIX, Nelly. **Kierkegaard, el único ante Dios**. Barcelona: Herder, 1977.

VOLTAIRE. **Dicionário filosófico.** In: Os pensadores. São Paulo: Abril Cultural, 1973.

TRADUÇÕES/TRANSLATIONS

O TRABALHO RECENTE SOBRE A CONTROVÉRSIA INTERNISMO- EXTERNISMO*

Laurence Bonjour

Tradução de Luiz Helvécio Marques Segundo¹

Embora o debate internismo-externismo tenha permanecido o principal foco da discussão epistemológica durante a última dúzia de anos ou mais, não é nada claro que algum progresso real rumo a uma solução tenha sido feito. Embora novos argumentos tenham sido aduzidos (a maioria em grande parte por Alvin Goldman: veja abaixo), tanto a definição das principais posições quanto a importância última da disputa parecem agora, se é que parecem, menos claros do que anteriormente pareciam. Como é aterradoramente frequente na filosofia, uma questão que inicialmente parecia clara e bem definida parece às vezes simplesmente se dissolver sob minucioso escrutínio.

Focar-me-ei, no espaço limitado que aqui tenho, em três tópicos centrais sob pena de nada dizer sobre muitas outras questões interessantes.² Começarei (§1) com uma discussão da principal distinção entre o internismo e o externismo. Depois (§2), discutirei alguns dos argumentos relevantes, focando-me principalmente na alegação recente de Goldman de ter demolido o internismo. Passo então (§3) à discussão da importância mais ampla da questão, tentando determinar em que medida as duas perspectivas são genuinamente rivais, e que escolha tem de ser feita.

§1. O que é o internismo?

Suporei aqui que o *internismo* é a perspectiva que principalmente precisa de clarificação, sendo o *externismo* definido simplesmente como a negação do internismo. A primeira pergunta que precisa ser respondida diz respeito à principal aplicação desse termo: parece claro que é aplicado a abordagens de algum conceito epistêmico, mas a qual primariamente? Aqui, as principais alternativas, aparentemente, são o próprio conceito de conhecimento, o conceito de garantia, e o conceito de justificação epistêmica.

* “Recent Work on the Internalism-Externalism Controversy”, in *A Companion to Epistemology*, eds. Jonathan Dancy, Ernest Sosa e Mathias Steup. Blackwell, 2010, pp. 33-43.

¹ Revisor: Luiz Maurício Menezes.

² Devo mencionar em particular a defesa naturalisticamente orientada do externismo de Hilary Kornblith em seu 2002. Para uma discussão crítica, veja Bonjour 2006.

Não há dúvida de que o rótulo “internismo” (ou o rótulo contrastante “externismo”) pode ser aplicado a abordagens ao conhecimento; mas parece claro que, se vamos usá-los para marcar uma distinção interessante aqui, isso tem de ser em virtude deles se ligarem a algum elemento ou ingrediente mais restrito do conhecimento, presumivelmente ou à garantia ou à justificação – sendo que os outros ingredientes plausíveis do conhecimento, nomeadamente a crença, a verdade e alguma condição anti-Gettier parecem cada um deles se enquadrar claramente ou no internismo (a crença) ou no externismo (os outros dois). Michael Bergman sugeriu que a noção de internismo deveria ser considerada como pertencendo primariamente à *garantia*, entendida (no sentido técnico introduzido por Plantinga³) como seja o que for que se adicione à crença verdadeira para produzir conhecimento;⁴ mas além dessa inclusão da condição anti-Gettier obviamente externa, essa noção de garantia me parece altamente artificial, correspondendo a nada no pensamento comum, e tornando os principais argumentos intuitivos difíceis ou impossíveis de se avaliar. Por essas razões, seguirei inicialmente a maior parte da bibliografia tomando a noção de internismo e da distinção resultante como se aplicando primariamente às abordagens da *justificação epistêmica*, aceita por enquanto com uma das exigências para o conhecimento (junto com a crença, a verdade, e a condição anti-Gettier) – embora eu vá (no §3) discutir algumas razões para se duvidar do quão satisfatória é de fato essa concepção.

O que significa, então, dizer que uma abordagem da justificação é internista em caráter? O que exatamente supomos como *interno* ao quê? Num respeitado manual introdutório, Mathias Steup oferece a seguinte caracterização:

O que torna uma abordagem da justificação internista é que ela impõe certa condição àqueles fatores que determinam se uma crença é justificada. [...] A condição requer [que tais fatores] sejam *internos à mente do sujeito* ou, pondo de outro modo, *acessível à reflexão*.⁵

Temos aqui aquelas que claramente são as duas principais alternativas nas discussões recentes sobre uma abordagem ao internismo,⁶ alternativas que parecem de fato distintas.⁷ De acordo

³ Veja Plantinga, 1993, p. 3.

⁴ Bergmann, 1997.

⁵ Steup, 1996, p. 84.

⁶ Abordagens próximas aparecem, *e.g.*, em Fumerton, 1995, pp. 60-66 (embora Fumerton também mencione duas outras versões de internismo); e em Conee e Feldman, 2004, p. 55.

⁷ Steup tenta reconciliá-las dizendo que “a expressão ‘acessível à reflexão’ pode ser tomada como uma elaboração de como a expressão ‘interno à mente’ tem de ser entendida” (Steup, 1996, p. 85), mas isso parece claramente insatisfatório por razões que emergirão mais tarde em minha discussão.

com a primeira, à qual Conee e Feldman (talvez seus principais proponentes) adotaram o rótulo “mentalismo”, uma abordagem é internista apenas no caso em que recorre apenas a coisas que são “internas à vida mental da pessoa”, a “estados, eventos e condições mentais ocorrentes e disposicionais”.⁸ De acordo com a segunda, que é mais comumente referida com “internismo acessibilista”, uma abordagem internista é aquela de acordo com a qual os “fatores” justificativos – todos eles, de acordo com a maioria das versões comuns – têm de ser “acessíveis” à pessoa, num sentido em que é preciso pelo menos alguma clarificação. Vale a pena notar que a base para a concepção mentalista do internismo parece ser, surpreendentemente, metafísica em caráter; ao passo que, em contraste, a base para a concepção internista acessibilista é reconhecidamente epistemológica.⁹

Qual dessas duas perspectivas melhor capta a ideia intuitiva central do internismo? Ao pensar sobre essa questão será útil ter em mãos uma amostra da perspectiva *externista* da justificação epistêmica para servir como ponto dialético, pois certamente é um bom teste para uma abordagem internista que ela possa dar um tratamento claro daquilo que supostamente falta às perspectivas externistas. A escolha óbvia para esse propósito, e de fato a única perspectiva externista específica que será discutida, é o *fiabilismo*: a perspectiva de que uma crença é epistemicamente justificada se resulta de um processo cognitivo que é (suficientemente) fiável em produzir crenças verdadeiras.¹⁰ Argumentarei que o mentalismo faz um trabalho insatisfatório em captar o contraste pretendido com tal perspectiva externista e, que, por essa razão, o internismo acessibilista é preferível.¹¹

Há dois aspectos dessa razão. A primeira é que há exemplos possíveis de estados e processos mentais que não parecem produzir o tipo de justificação internista que seria contrastado da maneira correta com o fiabilismo. Considere, por exemplo, a ideia bastante aceita de estados mentais *inconscientes*, estados que presumivelmente poderiam ser combinados com processos

⁸ Conee e Feldman, “Internalism Defended”, em Conee e Feldman, 2004, pp. 55-56. Na formulação preferida deles, a tese internista é a de que a justificação epistêmica sobrevém a itens desse tipo.

⁹ Embora me foque quase que inteiramente nessas duas concepções de internismo, elas não são as únicas a serem encontradas na bibliografia. É razoavelmente claro, porém, que são as mais proeminentes e geralmente aceitas.

¹⁰ O criador e principal proponente do fiabilismo é, certamente, Alvin Goldman. Veja *inter alia* Goldman, 1986.

¹¹ A razão inicial de Conee e Feldman para adotar o mentalismo como abordagem ao internismo é que “ele se aproxima bastante das distinções [entre internismo e externismo] na sua contraparte na filosofia da mente e na ética” (Conee e Feldman, 2004, p. 57): as posições “internistas” nessas outras áreas são teses da sobrevida, e a formulação preferida de Conee e Feldman do mentalismo, como notado acima, sustenta que o estatuto justificacional é sobreveniente aos estados mentais. Mas essa razão obviamente não tem uma tendência real de mostrar que a mentalismo assim entendido é o melhor modo de capturar a distinção epistemológica que nos interessa: que há um paralelo significativo entre a epistemologia e essas outras áreas não é obviamente algo que possa ser aceito antes que uma abordagem epistemologicamente adequada seja oferecida.

mentais inconscientes. Não há razão aparente pela qual esses estados ou processos mentais não poderiam de algum modo tornar provável a verdade de uma crença particular, e nem pela qual eles poderiam não contribuir causalmente para se sustentar tal crença. Esse seria um tipo de processo cognitivo fiável, um processo que difere daqueles (tais como os processos perceptuais mais comumente citados pelos fiabilistas) apenas por serem inteiramente “internos” à mente. Mas por que esse fato metafísico deveria ter qualquer importância epistemológica, dado que em ambos os casos a razão a favor da crença que está disponível na situação não é uma razão cuja pessoa em questão normalmente estará ciente? Essa é uma queixa internista comum ao fiabilismo, mas parece também se aplicar à justificação internista (de acordo com o mentalismo) anteriormente descrita. Um tipo diferente de exemplo, que trata do mesmo ponto, é uma justificação que apela a algum tipo de propriedade do estado mental de uma pessoa – talvez a coerência de todo o seu sistema de crenças – que seja demasiado complexa e multifacetada para a pessoa ser sempre capaz de apreender refletivamente que foi obtida. Novamente, embora uma crença que resulte da presença de tal propriedade pudesse ser fiavelmente causada, é difícil ver por que uma justificação que apela a esse tipo de situação inteiramente mental deveria contar como internista de modo que as justificações fiabilistas comuns não: uma vez mais, há uma diferença metafísica, mas uma diferença que parece não ter importância epistemológica óbvia.

O outro lado da moeda é a possibilidade de elementos *externos* à mente mas que são, não obstante, capazes de serem diretamente apreendidos de uma maneira que lhes permita desempenhar um papel justificativo, um papel que contraste com uma justificação fiabilista típica naquilo que intuitivamente parece ser o modo correto. As teorias da percepção realistas diretas sustentam que os objetos materiais são capazes de serem apreendidos diretamente de uma maneira que permite que a sua presença na percepção justifique as crenças correspondentes. E muitas abordagens racionalistas da justificação *a priori* sustentam que entidades como universais, entidades matemáticas e conexões lógicas podem ser direta ou intuitivamente apreendidas e, por isso, desempenham um papel na justificação das afirmações *a priori*. Em ambos os casos há entidades mentais envolvidas, mas a suposta justificação parece envolver essencialmente a relação de tais estados com essas entidades não-mentais, de modo que as entidades não-mentais desempenham um papel indispensável.¹² Certamente que ambas as perspectivas foram seriamente criticadas, e meu próprio ponto de vista é que a

¹² Estou em débito aqui com uma valorosa discussão em Fumerton, 1995, pp. 60-62.

primeira está errada. Mas na verdade nenhuma delas está tão obviamente errada a ponto de ser completamente ignorada de maneira razoável. O ponto é então que em cada caso a justificação, embora recorra a entidades *não-mentais*, parece contrastar com a justificação fiabilista exatamente na maneira correta epistemologicamente significativa: a base para a justificação é aquela em que a pessoa é capaz de (i) apreender refletidamente e (ii) avaliar criticamente, e são essas as características, concedo, que tornam plausível classificá-la como internista em caráter.¹³

Sugiro, então, que a abordagem correta do internismo é no fim das contas o internismo acessibilista: o que principalmente importa não é o estatuto metafísico de um elemento ou fator justificativo, mas ao invés a sua *disponibilidade* à pessoa como uma *razão* (ou a base para uma razão) para se aceitar uma crença particular, uma razão que ela esteja então em posição de avaliar criticamente – pois é a falta dessas características que fazem uma justificação fiabilista típica parecerem insatisfatórias de acordo com as intuições reconhecidamente internistas. Isso é dizer que o tipo relevante de internalidade é *ser interno à perspectiva cognitiva de primeira pessoa de alguém*, no qual ser interno à sua mente num sentido metafísico é claramente não suficiente e defensavelmente não necessário.

Esse modo de entender o internismo tem dois corolários. Um, que, com efeito, temos pressuposto até agora, mas que às vezes é rejeitado,¹⁴ é que apenas uma perspectiva de acordo com a qual *todos* os elementos exigidos para se fornecer uma razão cogente para a crença em questão são apropriadamente acessíveis pode satisfazer a intuição internista fundamental. Somente se for assim é que a pessoa genuinamente *tem* uma razão, como oposta a meramente parte de uma razão, para a crença, uma razão que pudesse parecer justificar refletidamente a crença e que estivesse disponível para a avaliação crítica. E uma perspectiva externista será, então, aquela que permite que algum elemento que seja essencial à cogência de uma razão justificativa esteja fora da, externo a, perspectiva cognitiva da pessoa. É o fato de que isso é claramente possível para o fiabilismo que o torna uma perspectiva externista.

¹³ Outro exemplo seria a posição de G. E. Moore de que os dados dos sentidos, embora objetos da experiência direta e, por isso, aparentemente capazes de contribuir para a justificação, são eles próprios não-mentais em caráter – uma tese que é surpreendentemente plausível em pelo menos algumas concepções dos dados dos sentidos e daquilo que conta com mental. Veja Moore, 1922, pp. 1-30.

¹⁴ E.g. por Bergmann, 1997.

O outro corolário, que foi ainda mais desafiado, é que a “acessibilidade” dos fatores ou elementos justificativos têm de ser entendidos de maneira bastante forte. Parte desse desafio já fora feito na insistência de Steup e outros sobre a acessibilidade *reflexiva*: ter acesso à referência útil ou a uma situação perceptível que pudesse fornecer uma razão não torna essas razões internamente acessíveis da maneira correta, uma vez que até que eu faça aquilo que é exigido para usá-las, não tenho *efetivamente* a razão em questão. Mas do mesmo modo (e eis outro problema com o mentalismo) o fato de que uma base para uma razão de algum modo escondida em minha experiência perceptiva, ou no sistema de crenças, ou em outros conteúdos mentais de um modo inteiramente desapercibido também não parece me fornecer uma razão apropriadamente internista para a crença correspondente (novamente, a coerência irreconhecida do meu sistema de crenças completo é um bom exemplo): até que eu faça aquilo que é necessário para isolar os ingredientes de tal razão e juntá-los de maneira apropriada, novamente não *tenho* efetivamente a razão em questão e, obviamente, não estou em posição de avaliá-la criticamente. Não quero aqui insistir que tudo isso tem de ser feito de maneira explícita e completamente manifesta, embora de um ponto de vista estritamente epistemológico isso seja obviamente ideal. Mas algum tipo de consciência tácita ou implícita daquilo que está envolvido é exigida se tenho que ter uma razão efetiva ao invés de uma razão meramente potencial. E uma razão meramente potencial não difere de uma justificação fiabilista na maneira correta: qualquer justificação fiabilista individual é uma justificação que as pessoas pelo menos em princípio poderiam se tornar explicitamente cientes dela, e o mero fato de que tal consciência é mais fácil de alcançar num caso do que no outro não parece em si fazer uma diferença epistemológica na justificação da crença na situação em que tal consciência ainda não foi alcançada.

§ 2. Argumentos recentes

Passando aos argumentos a favor e contra essa concepção de internismo e a concepção correlata de externismo, quero primeiro considerar alguns argumentos que se tornaram padrão em cada lado (focando-me novamente inteiramente nas versões fiabilistas do externismo), mas que pouco precisa ser dito num exame dos desenvolvimentos recentes. Do lado internista temos: (1) o apelo a casos do gênio maligno cartesiano e similares a fim de mostrar que a fiabilidade não é *necessária* para a justificação; (2) o apelo a exemplos de clarividência e similares a fim de mostrar que a fiabilidade não é suficiente para a justificação; e (3) a

objeção ao fiabilismo que deriva do problema de se escolher a especificação geral relevante para determinar a fiabilidade de um processo cognitivo específico (“o problema da generalidade”). Do lado externista temos: (1) o argumento de que sujeitos epistêmicos não sofisticados têm crenças justificadas e conhecimento, muito embora eles não satisfaçam as exigências de qualquer abordagem internista plausível; e (2) o argumento de que o internismo conduz inevitavelmente ao ceticismo.¹⁵ Fora os argumentos mais específicos de Goldman, considerados abaixo, alguns dos quais equivalem a elaborações mais específicas do argumento externista (2), tem havido relativamente pouca coisa nova na discussão desses argumentos no período aqui examinado. Um ponto que merece ser mencionado, porém, é que continua a não haver sequer uma aproximação à solução do problema da generalidade, o que fortemente me faz pensar que simplesmente não há solução a ser encontrada – e, por conseguinte, que a principal perspectiva externista, no fim da contas, não pode sequer ser formulada claramente.¹⁶

Focar-me-ei aqui quase que inteiramente naquela que certamente é a contribuição mais substancial ao panorama argumentativo nessa área no período relevante, a saber, a extensa e, na opinião do autor, completamente devastadora crítica de Goldman ao internismo em seu artigo “Internalism Exposed”.¹⁷ Nesse artigo, Goldman argumenta que tanto o internismo quanto os principais argumentos a seu favor, quando cuidadosamente examinados, estão “cheios de problemas”. Ele conclui: “Não vejo esperanças para o internismo; ele não sobrevive à ribalta” (p. 293).

Na verdade, há muitos aspectos da discussão de Goldman que são seriamente problemáticos. Embora ele comece com o que parece ser uma formulação do internismo acessibilista, a sua discussão, contudo, frequentemente parece pressupor o mentalismo, aparentemente sem se aperceber das diferenças entre essas perspectivas. A sua concepção da razão básica a favor do internismo em termos daquilo que ele chama de “concepção de orientação deontológica da justificação”, embora inegavelmente reflita algo que alguns internistas tenham dito, parece-me não captar a intuição internista central. E as versões específicas do internismo que ele

¹⁵ Para uma discussão mais detalhada desses argumentos, veja BonJour, 2002.

¹⁶ Para uma excelente discussão desse problema, veja Conee e Feldman, “The Generality Problem for Reliabilism”, em Conee e Feldman, 2004, pp. 135-165. A tentativa recente melhor desenvolvida e mais extensa de resolver o problema é de longe a de William Alston, no capítulo 6 de seu 2005. Contrário a alegação de Alston, não estou convencido de que ele tenha sido bem sucedido em resolver o problema, mas há pouca dúvida de que a sua discussão é o ponto de partida para qualquer tentativa futura nessa direção.

¹⁷ Goldman 1999. As referências entre parênteses nesta seção serão das páginas desse artigo.

formula ao longo do artigo não correspondem muito bem a algumas das idéias que os internistas reais sempre defenderam. Parece-me que tudo isso dá um caráter de certo modo espantalhesco a uma boa parte de sua discussão.¹⁸ Na presente discussão, porém, tratarei primariamente de alguns dos problemas mais específicos que Goldman levanta para o internismo. Não penso que quaisquer desses argumentos apóiem a conclusão de Goldman de que o internismo é fundamentalmente insustentável. Mas levantam problemas que uma perspectiva internista precisa tratar, problemas que podem de fato contribuir de maneira importante para refinar e clarificar a posição internista.

Considere primeiro aquilo que Goldman de “o problema das crenças estocadas”:

[...] Num dado momento a vasta maioria das crenças de uma pessoa estão estocadas na memória ao invés de serem ocorrentes ou estarem ativas [...] Ademais, para quase todas essas crenças o estado consciente dessa pessoa nesse momento nada inclui que as justifique. Nenhuma experiência perceptiva, nenhum evento de memória consciente, e nenhuma premissas conscientemente levadas em conta no momento selecionado serão justificacionalmente suficientes para tal crença. De acordo com o internismo forte, então, nenhuma dessas crenças está justificada naquele momento. (p. 278) (“Internismo forte” aqui é a perspectiva de que apenas fatos sobre os estados conscientes de um agente num momento particular podem justificar suas crenças naquele momento.)

De fato, o principal problema com o qual Goldman parece estar preocupado aqui, na verdade, não se limita como tal às crenças estocadas, mas tem a ver ao invés com o problema de se *num dado momento* pode haver conscientemente uma base adequada na mente que justifique praticamente qualquer crença que se possa escolher, esteja estocada ou não. Mas, embora alguns internistas tenham talvez aceitado tolamente a limitação ao que está disponível no momento (por mais extensa que seja!), não há, sugiro, qualquer coisa acerca da base internista que requeira de algum modo tal perspectiva quixotesca. O que está diretamente disponível a partir da perspectiva epistêmica em primeira pessoa não deixa de estar disponível ou se torna de algum modo externo em caráter só porque tem de ser reunido, revisto e examinado a todo momento. Na verdade, uma razão justificativa ou argumento cujos elementos foram reunidos ao longo do tempo têm ainda de ser apreendidos de algum modo como um todo unificado

¹⁸ Para alguma elaboração desses pontos, veja BonJour 2001. Esse artigo também contém uma discussão mais completa dos argumentos de Goldman do que é possível aqui, e é dele que tirei algum do material da presente discussão. (Para outras consideração internistas dos argumentos de Goldman bastante diferente, veja Earl Conee e Richard Feldman, “Internalism Defended”, em Conee e Feldman 2004, pp. 53-82).

com a ajuda da memória e talvez de registros escritos. Mas não há razão pela qual essa apreensão tenha de ser *momentânea*. Assim, a primeira objeção não é forte o bastante contra formas razoáveis de internismo.¹⁹

A segunda objeção que pretendo discutir é aquela a que Goldman chama “o problema dos indícios esquecidos”. Ela diz respeito aos casos em que a pessoa em questão simplesmente esqueceu-se da base indiciária sobre a qual uma crença foi originalmente aceita, mas ainda mantém tal crença. Assim, em seu exemplo, Sally leu, na seção de ciência do *New York Times* (que podemos aceitar como sendo uma fonte altamente fiável), um artigo sobre os benefícios à saúde de se comer brócolis e formou a crença correspondente. Ele ainda tem a crença, porém esqueceu-se como a adquiriu. Assim, ela parece não ter disponível uma justificação apropriadamente internista, embora, de acordo com Goldman, a sua crença ainda seja justificada (p. 280).²⁰

Na verdade, se Sally tem uma razão internista a favor de sua crença, e o quão forte ela é, dependerá de outros detalhes do caso, alguns deles bastante sutis. Uma questão é se Sally tem boas razões para pensar que ela é geralmente cuidadosa com as fontes pelas quais ela aceita suas crenças – ou talvez apenas as crenças sustentadas com o grau de segurança com o qual ela sustente essa crença. Outra questão é se Sally *acredita* que adquire a crença de uma fonte fiável, ainda que não possa se lembrar qual é, e se tem razões para pensar que tanto seus juízos sobre a fiabilidade das fontes quanto as suas memórias são fiáveis. Uma terceira questão é se Sally pode se lembrar de vários tipos de detalhes que ampliam e reforçam a crença em questão: os detalhes sobre os modos específicos em que o brócolis leva a uma boa saúde e sobre como ele produz esses efeitos, ainda que as memórias desses detalhes sejam também crenças cujas bases justificativas originais ela não se lembra. Sally poderia muito bem ter boas razões internistas para pensar que as crenças pelas quais ela pode se lembrar dos detalhes desses tipos muito provavelmente foram derivadas de uma fonte fiável e talvez

¹⁹ Para ser sincero, há um problema próximo para *algumas* posições internistas. Reflete-se naquilo que Goldman chama “o problema da recuperação simultânea”, que é, como ele sugere, particularmente um problema para as razões ou argumentos justificativos característicos do coerentismo holista: é duvidoso que uma razão que dependa da coerência de todo o sistema de crenças de alguém ou mesmo de um subsistema muito amplo seja capaz de ser reunida e compreendida adequadamente com a ajuda da memória e registros escritos.

²⁰ E, se verdadeiro, esse caso constitui uma instância de conhecimento. Não vejo razão para um internista negar que haja um sentido ou uso de “conhecimento” no qual a crença de Sally conte como conhecimento (muito provavelmente, junto de outros sentidos em que não conta, e ainda outros nos quais o resultado é, de longe, incerto). Às vezes os epistemólogos parecem falar do *conhecimento* como se fosse uma bandeira que se esteja permitido balançar, ou mesmo um emblema que se possa vestir, e não tenho objeções a Sally balançar a bandeira do conhecimento, se ele assim o quiser (ou a alguém mais que queira).

também muito provavelmente acuradamente lembradas. Com base em algumas, ou todas essas respostas, Sally poderia ter os recursos para uma razão internista a favor de uma crença principal, cuja força obviamente variaria com os detalhes. Admitidamente, essas justificações provavelmente não são tão fortes quando as que ela teria no momento em que lia o artigo original, mas esse parece exatamente o resultado correto e de modo algum implausível.

Uma alegada terceira objeção surge do “problema do intervalo da decisão doxástica”. Goldman argumenta que as supostas variedades de internismo que permitem o apelo apenas a estados mentais conscientes (sejam ocorrentes ou estocados na memória) têm de ser expandidas a fim de permitir o acesso às relações lógicas ou probabilísticas, e propõe que isso seja feito permitindo-se que algumas das “propriedades formais dos estados mentais, isto é, as propriedades formais e matemáticas de seus conteúdos” contem como parte da base para as justificações ou razões internistas, embora insista que deveriam se restringir àquelas razões que “são conhecíveis pelo agente no momento da decisão doxástica” através do emprego de uma gama de “relações computacionais ou algoritmos” (pp. 282-3). O problema é então quanto tempo deveria ser permitido a tais computações, um problema que é agravado, de acordo com Goldman, pela preocupação de que os estados mentais de um agente poderiam mudar durante o intervalo permitido de tal modo que afeta a justificação da proposição em questão (pp. 283-4).

Mas esse modo de formular o suposto problema me parece errado de diversos modos. Como já vimos, nada há sobre a base a favor do internismo que limite os fatores ou elementos justificativos relevantes aos estados mentais. Se as propriedades ou relações lógicas são diretamente acessíveis via raciocínio *a priori*, então podem também desempenhar um papel justificativo internista. Além do mais, embora essa seja uma questão mais discutível, não vejo qualquer razão para se limitar as propriedades e relações em questão àquelas que são *formais* num sentido interessante – ou, menos ainda, àquelas que são conhecidas via “procedimentos computacionais ou algoritmos”.²¹

Mais importante, a ideia de um “intervalo de decisão” fixo dentro do qual as propriedades lógicas e probabilísticas elegíveis têm de ser determinadas, embora ainda exclua a

²¹ As pessoas de fato raramente fazem uso de procedimentos desse tipo para decidir questões lógicas como opostas a questões matemáticas. Certamente a imagem que Goldman evoca a certa altura de um aspirante a coerentista tentando estabelecer a consistência de seu conjunto de crenças usando uma tabela de verdade muito longa não é uma perspectiva plausível das operações cognitivas reais de *alguém*.

possibilidade de uma mudança mental significativa, é artificial ao extremo. Obviamente que mudanças de vários tipos nos estados mentais de alguém podem afetar questões de justificação internista, mas isso, tanto quanto posso ver, não tem qualquer conexão essencial com a questão sobre as propriedades e relações conhecíveis *a priori* a que podemos recorrer. O comportamento de tais mudanças mentais na justificação é simplesmente uma questão independente, e não vejo razão pela qual as duas têm de ser tratadas conjuntamente da maneira que Goldman sugere.²²

Um quarto problema, e de certo modo relacionado, é “o problema da disponibilidade”. Embora Goldman apresente a questão em relação ao suposto conjunto de operações computacionais cujos resultados são elegíveis pela inclusão de justificações ou razões internistas (p. 285), penso que, por razões já suficientemente indicadas, é melhor apresentá-la simplesmente como o problema de quais propriedades ou relações discerníveis *a priori* são assim elegíveis. O que supostamente cria o problema são as grandes disparidades que existem entre as pessoas no que diz respeito à sua capacidade, como um resultado de treinamento e habilidade intelectual básica, de alcançar tais *insights a priori*. E o problema então é que *insights* são internisticamente admissíveis.

A minha resposta a esse suposto problema é simplesmente que não consigo ver por que há algum problema aqui afinal. Parece óbvio que a resposta deveria ser que um *insight* particular *a priori* pode desempenhar um papel nas justificações ou razões internistas de uma dada pessoa apenas no caso em que essa pessoa é capaz de compreendê-lo ou apreendê-lo. As relações lógicas que são demasiado complicadas ou sutis para eu aprender não podem contribuir para a *minha* justificação de minhas crenças, mas caso você possa apreendê-las, elas podem desempenhar perfeitamente bem um papel em *sua* justificação. Como diz Goldman, isso significa que “duas pessoas precisamente no mesmo estado indiciário (em termos de situação perceptiva, crenças de fundo, e assim por diante) poderiam ter permissões epistêmicas diferentes” (p. 286), isto é, poderiam diferir naquelas crenças que são justificadas

²² Poderia, contudo, parecer haver uma questão de quanto tempo uma pessoa levaria para alcançar *insights a priori* de vários tipos ou mesmo argumentos justificativos resultantes, mas é difícil ver a importância disso. O internista diria simplesmente que a justificação pode resultar apenas daquelas conexões lógicas ou probabilísticas que foram efetivamente reconhecidas – embora tenhamos de adicionar que tal reconhecimento possa ser mais ou menos preciso e explícito, podendo a força da razão resultante variar de acordo com isso. Se alguém demora mais e, com isso, alcança uma cogente, embora complicada, então, na medida em que essa razão é adequadamente apreendida por fim, não há razão aparente pela qual um internista devesse excluí-la com base no tempo que levou.

para eles. Mas é difícil ver por que alguém deveria achar isso surpreendente ou objetável de todo.

A quinta e última das questões que pretendo considerar aqui tem a ver com a acessibilidade aos princípios epistêmicos que subjazem às supostas razões internistas a agentes epistêmicos ingênuos ou comuns. Goldman argumenta, embora não exatamente desse modo, que tais princípios deveriam ser vistos como partes essenciais das razões em questão, de modo que eles teriam de ser acessíveis, presumivelmente em bases *a priori*, a partir da perspectiva epistêmica em primeira pessoa se tais razões têm de ser internisticamente aceitáveis. Inclino-me a concordar. E o seu argumento adicional de acordo com essa base é que a maioria ou todos os agentes epistêmicos, e de fato, pelo menos alguns epistemólogos, não são capazes de formular e reconhecer tais princípios. Do modo como ele formula a tese básica internista, isso significa que tais princípios não são elegíveis como componentes de razões ou justificações internistas a qualquer um, conduzindo assim ao “completo ceticismo” (pp. 287-8).

Esse é o mais sério dos problemas que Goldman levanta e o mais difícil de tratar num curto espaço. Mas podemos ver imediatamente que algo está errado com a sua formulação da questão se perguntarmos por que um princípio epistêmico que seja genuinamente auto-evidente de um ponto de vista *a priori* para um agente epistêmico deveria ser considerado inelegível para contribuir para a razão ou justificação da crença do agente só porque acontece dele ser demasiado complicado ou sutil de ser discernível por outros agentes epistêmicos menos sofisticados. Certamente que a coisa certa a se dizer aqui é que os princípios epistêmicos que contribuem para as razões ou justificações de um agente têm de estar disponíveis a qualquer momento é necessário *àquele agente*. Isso certamente significará que a acessibilidade de tais razões variará de um agente a outro, embora eu não consiga ver que haja algo de implausível sobre tal resultado.

A segunda coisa a dizer sobre esse problema é que a disponibilidade pode ser uma questão de grau, não exigindo a capacidade de formulação explícita de todos os graus de disponibilidade. Aqui a situação é bastante paralela à situação dos princípios lógicos. Os agentes ingênuos comuns provavelmente não conseguem formular explicitamente um princípio como o *modus ponens*, mas podem não obstante ser capazes, após sua formulação e explicação, de reconhecê-lo como o princípio que estavam seguindo. E algo análogo pode muito bem acontecer com vários princípios epistêmicos, embora o grau ao qual um princípio será

plausível variará amplamente de caso para caso. Tanto esse ponto quanto o anterior significarão que o grau ao qual várias crenças são justificadas de um ponto de vista internista provavelmente variará de pessoa a pessoa, mas uma vez que a questão de quem balança a bandeira do conhecimento é posta de lado como desinteressante, como acredito que deva ser, é difícil ver que há algo de alarmante ou mesmo particularmente surpreendente em tal resultado.

A última coisa que quero dizer nessa ocasião sobre essa última questão é que, embora seja relevante para a avaliação das perspectivas internistas particulares, ela não constitui qualquer objeção real ao internismo. Se um internista chega a uma abordagem de outra forma plausível da justificação de um tipo particular de crença, mas que depende de um princípio que não seja plausivelmente um princípio que as pessoas comuns sequer estejam cientes, então essa abordagem da justificação não terá qualquer relevância ao problema de se a crença ou crenças em questão são justificadas a eles – embora pudesse ainda ser de grande interesse enquanto abordagem de como a crença ou crenças em questão poderiam estar justificadas para aqueles que venham a reconhecer o princípio em questão. Não penso que Goldman esteja correto de que todas as abordagens internistas da justificação dos principais tipos de crenças que o senso comum considera como justificadas ou razoáveis terão de rejeitar esse estatuto; mas, caso o façam, esse seria simplesmente um resultado filosófico a ser respeitado como qualquer outro. Se tal resultado parece implausível, como acredito que seja, é porque, concedo, todos nós acreditamos que temos boas razões para as nossas crenças sobre o mundo, e não porque acreditamos que estejam justificadas de uma maneira não específica que não precise envolver a posse de tais razões.

§3- Há uma controvérsia genuína?

Tendo dito tudo isso em explanação da posição internista e em sua defesa contra as objeções de Goldman, darei um pequeno passo atrás e perguntarei se a disputa entre internismo e externismo é de fato tão claramente definida quanto ainda muito comumente se considera. Uma razão para se duvidar se isso é assim é o caráter aparentemente intratável da disputa, pouco comum até para uma questão filosófica, na qual as partes opostas não apenas não conseguem chegar a um acordo mas frequentemente tem bastante dificuldade em concordar até sobre aquilo que conta como uma boa razão para um lado ou para o outro.

Outra razão é a maneira pela qual a controvérsia é comumente formulada: como uma disputa sobre a abordagem ou análise correta da suposta propriedade da *justificação epistêmica*. Mas que propriedade exatamente é essa? O termo é obviamente um termo técnico, não claramente presente no senso comum, e vale a pena perguntar se há algum modo de se especificar o seu suposto significado que seja neutro o bastante aos dois lados opostos para fornecer um alvo claro para a disputa. A resposta mais comum aqui é aquela brevemente já aludida anteriormente. A justificação epistêmica é supostamente uma das exigências para o *conhecimento*: a que precisa ser adicionada à crença, à verdade, e à satisfação de uma condição anti-Gettier.

Mas esse modo de se especificar o conceito alvo funcionará apenas se houver um conceito claro e inequívoco do próprio conhecimento, algo que me parece cada vez mais duvidoso. Alguns epistemólogos têm de fato sugerido que há pelo menos dois conceitos bastantes diferentes de conhecimento, um pelo menos predominantemente externista em caráter e um predominantemente internista em caráter,²³ uma sugestão que me parece plausível, mas ainda bastante restrita. E se algo assim é o caso, então um apelo ao conceito de conhecimento obviamente não servirá para apanhar o conceito único de justificação epistêmica sobre o qual internistas e externistas supostamente estão em desacordo.

De fato, dado o caráter intelectual do termo “justificação epistêmica”, não poderia ser o caso de simplesmente haver conceitos diferentes e incomensuráveis de justificação epistêmica, um (ou mais?) deles internista e um (ou mais?) deles externista – deixando pouco claro em que sentido eles competem numa escolha? Mesmo um internista de carteirinha com eu pode reconhecer que há questões importantes e claramente epistemológicas para as quais uma abordagem predominantemente externista pareça completamente apropriada e talvez mesmo preferível. A maioria dessas questões enquadra-se naquilo que Philip Kitcher rotulou apropriadamente “o projeto epistemológico aperfeiçoador”,²⁴ isto é, o projeto geral de avaliação e aperfeiçoamento da fiabilidade dos esforços cognitivos humanos num sentido amplamente empírico.²⁵ Nem me parece que qualquer externista razoável devesse se indispor

²³ Para versões de tal perspectiva, veja Mackie 1976, pp. 217-220, e Sosa 1991, p. 240 (e alhures no mesmo volume).

²⁴ Em Kitcher 1992, pp. 64-65.

²⁵ Goldman, no artigo discutido acima, descreve em algum detalhe uma investigação desse tipo geral: uma investigação psicológica das diferenças fenomenológicas (embora ele não use esse termo) entre memórias aparentes de percepções que genuinamente refletem percepções prévias e aquelas que são meramente produtos da imaginação, sendo a primeira obviamente mais fiável no que diz respeito à verdade das afirmações resultantes

a reconhecer o valor e a importância dos tipos de considerações epistêmicas pelas quais se interessa o internista.

A versão mais explícita e desenvolvida do tipo geral de perspectiva que estou sugerindo aqui é a defendida por Alston em seu livro recente *Beyond "Justification"*.²⁶ Embora Alston certamente não negue que os epistemólogos pudessem estipular vários sentidos para o termo "justificação", criando com isso conceitos de justificação que pudessem talvez ser úteis a vários propósitos, ele nega completamente que (fora tais estipulações) haja qualquer "estatuto ou propriedade objetivos das crenças apanhados por 'justificado'." (p. 27). E a implicação, uma vez que internistas e externistas claramente não estão disputando o conteúdo de conceitos de justificação meramente estipulados, é que não há na verdade qualquer questão genuína em disputa entre eles. A alternativa de Alston ao que ele chama de abordagem dos "desiderata epistêmicos" à epistemologia, na qual há muitas qualidades epistêmicas diferentes que uma crença pode ter, incluindo aproximadamente pelo menos aquelas refletidas em ambas as perspectivas, internista e externista. Todos esses "desiderata" são epistemicamente valiosos; e, embora haja conexões interessantes entre eles, não há um objetivo real ao se argumentar a favor de um foco exclusivo ou mesmo primário de alguns deles como opostos aos outros. Se Alston estiver correto, a disputa internismo-externismo, em nome da qual muita tinta tem sido desperdiçada, evapora-se completamente.

Embora não haja espaço aqui para considerar os detalhes de seu argumento, a minha própria opinião é que Alston está de fato correto em grande parte: primeiro, em sua alegação de que simplesmente não há uma concepção unívoca, não estipulativa, de justificação epistêmica que seja objeto de disputa; segundo, em sua sugestão de que há muitos valores epistêmicos diferentes dentre os quais não há necessidade de escolher. Em particular, há muitos tipos diferentes de questões epistemológicas, incluindo muitas que naturalmente são abordadas a

sobre as coisas que alegadamente foram percebidas. As memórias genuínas de percepção prévia mostram-se, como foi reportado, mais ricas em informação sobre as propriedades percebidas como cor e som, mais ricas em informação contextual sobre o instante e o lugar em questão, e mais detalhadas, ao passo que as memórias espúrias que resultam da imaginação tendem a ser mais empobrecidas nesses aspectos, embora contenham muito mais informação sobre as operações cognitivas da pessoa em questão (pp. 290-291). Temos aqui uma investigação de uma perspectiva externa, em terceira pessoa, que produz resultados que poderiam claramente ser valiosos, por exemplo, na avaliação da fiabilidade do testemunho em vários contextos e que uma pessoa que tivesse familiaridade com eles poderia também aplicar na avaliação de suas próprias memórias aparentes. E *poderia* talvez se útil, embora dificilmente essencial, na formulação desses resultados dizendo que as crenças mnemônicas que satisfazem os critérios para serem genuínas são justificadas num sentido externista, ou mais especificamente, fiabilista.

²⁶ Alston 2005. As referências entre parêntesis nesta seção são das páginas desse livro.

partir de um ponto de vista da terceira pessoa, externista, junto com algumas que são mais naturalmente vistas como questões internistas. O que é mais infeliz dessa perspectiva é a tendência de cada lado da disputa internista-externista de reivindicar a posse exclusiva do campo da epistemologia.²⁷ Repetindo Alston (e, certamente, o Presidente Mao), por que não deixar as flores epistemológicas desabrocharem?

Tendo sido reconciliador nesse ponto, quero, porém, insistir que há um sentido claro no qual uma abordagem internista, além de ser intelectualmente legítima por si própria, tem um tipo fundamental de prioridade para a epistemologia como um todo. (É por isso que disse apenas que Alston estava “em grande parte” correto). Não importa quanto trabalho possa ser feito no delineamento das concepções externistas do conhecimento, da justificação ou da fiabilidade e na investigação de como eles se aplicam a vários tipos de crenças ou áreas de investigação, há um sentido no qual todos esses resultados são meramente hipotéticos e inseguros na medida em que não podem ser alcançados a partir de recursos disponíveis de uma perspectiva epistêmica em primeira pessoa. Se, por exemplo, um epistemólogo afirma que certa crença ou conjunto de crenças, sejam suas ou de alguém mais, foi alcançado de maneira fiável, mas o diz com base em seus próprios processos cognitivos cuja fiabilidade é para ele meramente um fato externo ao qual ele não tem acesso em primeira pessoa, então a sua conclusão apropriada é meramente a de que a crença ou crenças originalmente em questão são alcançadas de maneira fiável (e talvez por isso são justificadas ou constituem conhecimento nos sentidos externista) *se* os próprios processos cognitivos do epistemólogo são fiáveis no sentido de que ele acredita que são. A única maneira aparente de alcançar um resultado que não seja em última instância hipotético nesse sentido é a fiabilidade de pelo menos alguns processos poderem ser estabelecidos com base naquilo que o epistemólogo pode saber direta ou imediatamente de sua perspectiva epistêmica em primeira pessoa. (O problema do regresso epistêmico nos espreita aqui).

E é por isso que o internismo é indispensável à epistemologia como um todo, ainda que não seja a melhor maneira de se abordar todos os problemas epistemológicos. Embora haja muitos outros problemas e questões legítimos, apenas uma abordagem internista funcionará em última instância quando, parafraseando agora o Bispo Butler, “Eu descansar num período de frescor” e indagar se em última instância tenho quaisquer boas razões para pensar que minhas

²⁷ Uma tendência que me parece (embora talvez eu esteja sendo tendencioso) ter sido manifestada mais fortemente por aqueles de persuasão externista (junto daqueles de persuasão naturalista).

crenças são verdadeiras – ou se de fato tenho quaisquer boas razões para pensar que são obtidas de maneira fiável.

Bibliografia

- Alston, W. (2005) *Beyond “Justification”: Dimensions of Epistemic Evaluation* (Ithaca, NY: Cornell University Press).
- Bergmann, M. (1997) “Internalism, Externalism and the No-defeater Condition,” *Synthese* 110, 399–417.
- BonJour, L. (2001) “The Indispensability of Internalism,” *Philosophical Topics* 29, 47–65.
- BonJour, L. (2002) “Internalism and Externalism,” in *The Oxford Handbook of Epistemology*, ed. P. Moser (Oxford: Oxford University Press), 234–63.
- BonJour, L. (2006) “Kornblith on Knowledge and Epistemology,” *Philosophical Studies* 127, 317–35.
- Conee, E. and Feldman, R. (2004) *Evidentialism* (Oxford: Oxford University Press).
- Fumerton, R. (1995) *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield).
- Goldman, A. (1986) *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
- Goldman, A. (1999) “Internalism Exposed,” *The Journal of Philosophy* 96, 271–93.
- Kitcher, P. (1992) “The Naturalists Return,” *Philosophical Review* 101.
- Kornblith, H. (2002) *Knowledge and Its Place in Nature* (Oxford: Oxford University Press).
- Mackie, J. L. (1976) *Problems from Locke* (Oxford: Oxford University Press).
- Moore, G. E. (1922) *Philosophical Studies* (London: Routledge and Kegan Paul).
- Plantinga, A. (1993) *Warrant: The Current Debate* (Oxford: Oxford University Press).
- Sosa, E. (1991) *Knowledge in Perspective* (Cambridge: Cambridge University Press).
- Steup, M. (1996) *An Introduction to Contemporary Epistemology* (Upper Saddle River, NJ: Prentice Hall).

POR QUE NÃO HÁ MAIS PROGRESSO NA FILOSOFIA?

David J. Chalmers

Tradução de Gregory Gaboardi¹

Há progresso na filosofia? Tenho duas reações a essa pergunta. A primeira é que a resposta é obviamente “sim”. A segunda é que essa é a pergunta errada. A pergunta certa não é “Há progresso?”, mas “Por que não há mais?”.

Podemos distinguir três perguntas sobre o progresso filosófico. A Pergunta da Existência: há progresso na filosofia? A Pergunta da Comparação: há tanto progresso na filosofia quanto na ciência? A Pergunta da Explicação (que tende a pressupor uma resposta negativa para pelo menos uma dessas duas perguntas): por que não há mais progresso na filosofia?

O que podemos chamar de uma visão “copo-meio-cheio” do progresso filosófico é a de que há algum progresso na filosofia. A visão copo-meio-vazio é a de que não há tanto quanto gostaríamos. Com efeito, a visão copo-meio-cheio consiste em uma resposta positiva para a Pergunta da Existência, enquanto a visão copo-meio-vazio (ou ao menos uma versão saliente dela) consiste em uma resposta negativa para a Pergunta da Comparação. Essas visões caem entre os extremos de uma visão copo-vazio que responde “não” para a Pergunta da Existência, dizendo que não há progresso na filosofia, e uma tese copo-cheio que responde “sim” para a Pergunta da Comparação, dizendo que há tanto progresso na filosofia quanto na ciência (ou tanto quanto gostaríamos).

Evidentemente a tese copo-meio-cheio e a tese copo-meio-vazio são consistentes entre si. Penso que para quase todos que se envolvem profundamente com a prática da filosofia ambas soarão verdadeiras. Em discussões sobre o progresso da filosofia minha experiência é que a maioria das pessoas se concentra na Pergunta da Existência: pessimistas sobre o progresso filosófico (e.g. Dietrich 2011, Nielsen 1987; McGinn 1993) defendem a tese do copo-vazio, e otimistas (e.g. Stoljar no prelo) respondem defendendo a tese do copo-cheio. Em vez disso, me concentrarei nas perguntas da Comparação e da Explicação. Articularei uma versão da tese do copo-meio-vazio, defendê-la-ei, e então abordarei a questão crucial do que a explica.

¹ Mestrando em Filosofia da PUCRS. Revisor: Rodrigo Reis Lastra Cid.

Devo dizer que este artigo é tanto um exercício em sociologia da filosofia quanto em filosofia. Na maior parte dele abstraí de minhas próprias posições filosóficas e metafisológicas para ter uma “visão de fora” do progresso filosófico, de uma perspectiva sociológica. Em grande parte do artigo estou amplamente dizendo o óbvio, mas algumas vezes vale dizer o óbvio para que o menos óbvio possa ser dito a partir dele. Somente perto do final trarei minhas próprias posições, que se inclinam um pouco mais em direção ao otimismo, e verei como a questão do progresso filosófico fica sob a luz delas.

1. A TESE CENTRAL

A forma de uma tese do copo-meio-vazio é: há menos progresso na filosofia do que em algum referencial. Para articular tal tese mais precisamente é necessário articular uma medida de progresso e um referencial. A medida de progresso que usarei é a convergência coletiva à verdade. O referencial que usarei é a comparação com as ciências duras.

Aqui tomo inspiração de Peter van Inwagen (2004, p.332), que escreve:

“A discordância na filosofia é disseminada e insolúvel. Não há quase nenhuma tese sobre a qual os filósofos concordam. Se há alguma tese que todos ou a maioria dos filósofos afirma, é uma tese negativa: que o formalismo não é a filosofia da matemática correta, por exemplo, ou que o conhecimento não é (simplesmente) crença verdadeira e justificada.

Não é assim que as coisas são nas ciências físicas. Concedo que a ‘parte avançada’ da física de partículas-elementares parece muito com a filosofia em questão de discordância fundamental e disseminada entre seus praticantes respeitados. Mas, na física há um grande corpo de teorias estabelecidas, usáveis e incontroversas, e de medições sabidamente precisas dentro dos limites que foram especificados. A ‘parte avançada’ da filosofia, entretanto, é basicamente toda ela.”

A tese de van Inwagen não é explicitamente sobre progresso, e o teor geral da discussão dele sugere algo mais próximo de uma tese do copo-vazio do que de uma tese do copo-meio-vazio. Penso que uma vez que o problema é um pouco melhor especificado, porém, a tese do copo-meio-vazio é mais defensável.

Aqui está minha tese central: não tem havido ampla convergência coletiva à verdade sobre as grandes questões da filosofia.

As grandes questões da filosofia aqui são questões como: qual é a relação entre mente e corpo? Como conhecemos o mundo exterior? Quais são os princípios fundamentais da moralidade? Há um deus? Temos livre arbítrio? Não tentarei oferecer uma lista mais precisa que essa, mas qualquer filósofo pode facilmente aparecer com uma lista de cerca de 10 grandes questões, e suspeito que haveria muita sobreposição entre essas listas. Poderíamos até usar essas listas para definir operacionalmente as grandes questões: as grandes questões de um campo no tempo t são aquelas que os membros desse campo contariam como as grandes questões do campo no tempo t . Para propósitos de comparação, podemos querer impor alguma regimentação na forma das grandes questões, por exemplo, formulando todas como escolhas entre um pequeno número de opções mutuamente excludentes.

Podemos definir a convergência coletiva sobre uma resposta em um período de tempo como o aumento no grau de acordo sobre a resposta desde o começo até o fim do período. O grau de acordo pode ser definido usando-se uma das várias medidas matemáticas para o acordo dentro de um grupo de pessoas sobre um conjunto de problemas.² A convergência coletiva (*simpliciter*) em um período de tempo é definida como a convergência coletiva sobre a resposta dominante no fim do período. O grau de acordo em um período sobre respostas para as grandes questões no mesmo período normalmente será baixo, conforme o desacordo sobre a resposta para uma questão é altamente relevante para torná-la uma grande questão. A convergência durante um período fornece uma medida mais discriminativa, que se pode razoavelmente esperar que seja alta em alguns casos e baixa em outros. Para nossos propósitos poderíamos escolher um período arbitrário (digamos, de 200 anos atrás até hoje),

² Deixarei aberta a questão de qual medida de acordo é a melhor para os presentes propósitos. Uma medida útil é a alfa de Krippendorff (Krippendorff 2013, pp.221-50), igual a $1-(D_o/D_e)$, onde D_o é a incidência observada de discordância entre os respondentes (somando todos os pares de respondentes e todas as questões) e D_e é a incidência esperada por puro acaso. Essa medida pode ser aplicada para comunidades de tamanhos diferentes (nas quais nem todos os membros precisam ter uma posição sobre o problema dado) e para questões cujas respostas têm muitos tipos diferentes de estruturas. A discordância é pesada por uma medida de “distância” entre duas respostas quaisquer, o que torna alfa particularmente útil para comparar questões com números diferentes de respostas. Com tal métrica em mãos, pode-se usar uma versão de alfa para medir o grau de acordo comunal a uma resposta específica. Para nossos propósitos algum ajuste de escala pode ser útil (e.g. impor um limite inferior de zero e então alinhar).

ou talvez melhor, escolher muitos períodos diferentes e obter a convergência média durante esses períodos.³

Podemos dizer que a *ampla* convergência coletiva durante um período exige tanta convergência quanto houver sobre as grandes questões nas ciências duras no mesmo período. Aqui assumirei as ciências duras como incluindo ao menos a matemática e as ciências naturais: paradigmaticamente física, química e biologia. Deixarei de lado as ciências cognitivas e sociais, que alegavelmente viram menos convergência nas respostas para suas grandes questões. Para os presentes propósitos não preciso tomar partido sobre como a filosofia fica relativamente à estas.

Ampla convergência coletiva à verdade em um período requer ampla convergência coletiva às respostas *verdadeiras* para as grandes questões do período. Isto é, precisamos ter tanto aumento de acordo sobre as respostas verdadeiras para as grandes questões quanto há nas ciências duras. Por causa da referência à verdade, a ampla convergência coletiva requer um grau de realismo sobre os domínios em questão. Porém, algo como a convergência à verdade é necessário para que a convergência constitua progresso e não retrocesso.

2. ARGUMENTO PARA A TESE CENTRAL

Aqui está um argumento para a tese central. Ele tem duas premissas, uma premissa empírica e uma premissa conectora.

- (1) Premissa empírica: Não tem havido ampla convergência coletiva sobre as grandes questões da filosofia.
- (2) Premissa conectora: Se não tem havido ampla convergência coletiva sobre as grandes questões da filosofia, então não tem havido ampla convergência coletiva à verdade sobre as grandes questões da filosofia.

³ Aqui não estou invocando a noção matemática padrão de convergência, que se aplica para séries infinitas ou períodos infinitos em vez de períodos discretos, e que requer que a comunidade eventualmente chegue arbitrariamente perto do consenso universal sobre um problema. Mas, um elemento de seu tempero estará presente se nossa medida de concordância tiver algum viés direcionado à concordância universal (como a versão de escala ajustada do alfa de Krippendorff tem), de modo que, por exemplo, uma mudança de 79% para 99% da comunidade concordando sobre uma resposta para uma questão binária conta como uma convergência maior do que uma mudança de 40% para 60%.

(3) Conclusão: Não tem havido ampla convergência coletiva à verdade sobre as grandes questões da filosofia.

O argumento é válido. A premissa conectora pode parecer uma verdade lógica, mas não é. A antecedente dessa premissa condicional diz que há menos convergência sobre respostas dominantes para as grandes questões na filosofia do que nas ciências duras, enquanto a conseqüente diz que há menos convergência sobre respostas verdadeiras para essas questões. Elas podem se separar de modo que a condicional se torna falsa se houver forte convergência ao falso (ou convergência forte não-correlacionada com a verdade) nas ciências junto com fraca convergência à verdade na filosofia. Ainda assim, dado que a convergência na ciência é amplamente convergência à verdade, ou meramente que seja tão provável que a convergência na ciência seja convergência à verdade quanto a convergência na filosofia, a premissa é plausível.⁴

Alguém poderia se preocupar que porque a maioria das teorias científicas eventualmente se mostra falsa, a maioria da convergência na ciência será convergência ao falso. Essa preocupação nos pressiona menos se formularmos as questões em termos de um pequeno número de escolhas mutuamente excludentes, como sugerido antes. Desse modo a convergência à verdade vai exigir apenas convergência para uma classe pouco-individuada e correta de teorias, tornando bem mais plausível que muita convergência na ciência tenha sido convergência à verdade. Alternativamente, se as questões permitem um âmbito aberto de respostas, então invocar uma métrica para a distância entre respostas (como também sugerido antes) permitirá que o acordo sobre teorias falsas que, falando estritamente, sejam relativamente próximas à verdade, constitua um tipo de convergência à verdade.

O trabalho principal no argumento é feito pela premissa empírica. Assumo que ela será plausível para os que têm uma breve familiaridade com a prática filosófica e científica. Ainda assim, ela faz uma alegação sociológica e não pode ser decisivamente estabelecida da poltrona. Nós não temos todos os dados empíricos necessários para uma investigação sistemática da premissa, mas temos um pouco deles.

A Pesquisa PhilPapers de 2009 (Bourget e Chalmers, 2014) interrogou filósofos profissionais sobre respostas para trinta questões importantes na filosofia. A pesquisa foi

⁴ Obrigado a Hedda Hessel Morch e Rory Madden por apontarem maneiras em que a premissa conectora poderia se mostrar falsa.

enviada para membros de 99 dos principais departamentos de filosofia (na maioria especializados em filosofia analítica/anglocêntrica) na América do Norte, Europa e Australásia. Cerca de 47% dos 2000 que receberam a pesquisa preencheram-na e retornaram-na. As questões foram colocadas como escolhas entre duas, três ou quatro opções. Os respondentes poderiam indicar que eles “aceitam” ou “se inclinam” para uma opção, ou dar uma variedade de respostas “outras” (e.g. sem familiaridade com o assunto, a questão é muito ambígua para ser respondida, não há fato que resolva a questão, aceito outra opção, e por aí vai). Os resultados (misturando as respostas “aceito” e “se inclinam”, e misturando as respostas “outras”) foram como segue:

- (1) Conhecimento a priori: sim 71%, não 18%, outros 11%.
- (2) Objetos abstratos: platonismo 39%, nominalismo 38%, outros 23%.
- (3) Valor estético: objetivo 41%, subjetivo 35%, outros 24%.
- (4) Distinção analítico/sintético: sim 65%, não 27%, outros 8%.
- (5) Justificação epistêmica: externismo 43%, internismo 26%, outros 31%.
- (6) Mundo exterior: realismo não-cético 82%, ceticismo 5%, idealismo 4%, outros 9%.
- (7) Livre arbítrio: compatibilismo 59%, libertarismo 14%, sem livre arbítrio 12%, outros 15%.
- (8) Deus: ateísmo 73%, teísmo 15%, outros 13%.
- (9) Alegações de conhecimento: contextualismo 40%, invariancismo 31%, relativismo 3%, outros 26%.
- (10) Conhecimento: empirismo 35%, racionalismo 28%, outros 37%.
- (11) Leis da natureza: não-humeano 57%, humeano 25%, outros 18%.
- (12) Lógica: clássica 52%, não-clássica 15%, outros 33%.
- (13) Conteúdo mental: externismo 51%, internismo 20%, outros 29%.
- (14) Meta-ética: realismo moral 56%, antirrealismo moral 28%, outros 16%.
- (15) Metafilosofia: naturalismo 50%, não-naturalismo 26%, outros 24%.
- (16) Mente: fisicismo 57%, não-fisicismo 27%, outros 16%.
- (17) Julgamento moral: cognitivismo 66%, não-cognitivismo 17%, outros 17%.
- (18) Motivação moral: internismo 35%, externismo 30%, outros 35%.
- (19) Problema de Newcomb: duas caixas 31%, uma caixa 21%, outros 47%.
- (20) Ética normativa: deontologia 26%, consequencialismo 24%, ética das virtudes 18%, outros 32%.

- (21) Experiência perceptual: representacionismo 32%, teoria dos qualia 12%, disjuntivismo 11%, teoria dos dados dos sentidos 3%, outros 42%.
- (22) Identidade pessoal: visão psicológica 34%, visão biológica 17%, visão do fato-adicional 12%, outros 37%.
- (23) Política: igualitarismo 35%, comunitarismo 14%, liberalismo 10%, outros 41%.
- (24) Nomes próprios: Millianos 34%, Fregeanos 29%, outros 37%.
- (25) Ciência: realismo científico 75%, antirrealismo científico 12%, outros 13%.
- (26) Teletransporte: sobrevivência 36%, morte 31%, outros 33%.
- (27) Tempo: Teoria B 26%, Teoria A 16%, outros 58%.
- (28) Problema do bondinho: puxa 68%, não puxa 8%, outros 24%.
- (29) Verdade: correspondência 51%, deflacionária 25%, epistêmica 7%, outra 17%.
- (30) Zumbis: concebíveis mas não metafisicamente possíveis 36%, metafisicamente possíveis 23%, inconcebíveis 16%, outros 25%.

O grau de discordância aqui é impactante, ainda que previsível. Apenas uma posição (realismo não-cético sobre o mundo exterior) atrai mais de 80% de adesão. Três posições (conhecimento a priori, ateísmo, realismo científico) atraem mais de 70% de apoio, com dissidência significativa, e outras três posições atraem mais de 60% de apoio. Nas outras 23 questões a posição dominante tem menos de 60% de apoio.

Aceitavelmente nem todas as questões estão entre as “grandes questões” do passado, mas certamente algumas estão: as questões sobre o mundo exterior, livre arbítrio, Deus, conhecimento, meta-ética, metafilosofia, mente e ética normativa, por exemplo. Apenas duas dessas (mundo exterior, Deus) têm posições com mais de 60% de apoio (e no caso da questão do mundo exterior, o consenso é um tanto enganoso, pois plausivelmente a maior questão é *como* conhecemos o mundo exterior).

Para que dados mais completos nos ajudassem a julgar a tese central precisaríamos dos resultados da pesquisa PhilPapers não apenas de 2009, mas em intervalos regulares no passado: 1909, 1809, e por aí vai. Em cada ponto teríamos que perguntar aos membros da comunidade filosófica: primeiro, quais questões eles aceitam como as grandes questões da filosofia, e segundo, quais respostas eles tomam para tais questões, bem como para as grandes questões de pesquisas anteriores. Também precisaríamos ter pesquisas longitudinais análogas em outros campos: a pesquisa MathPapers, a pesquisa PhysPapers, a pesquisa ChemPapers, a

pesquisa BioPapers, e por aí adiante. E precisaríamos de uma medida razoável de acordo por período. Prevejo que se tivéssemos tais pesquisas e medidas encontraríamos muito menos convergência nas respostas para as grandes questões sugeridas por pesquisas anteriores dos filósofos do que encontraríamos nas respostas correspondentes em outros campos.

Alguns dados parciais são fornecidos pelos 23 problemas que David Hilbert colocou para a matemática em 1900 (Hilbert, 1902; Yandell, 2002). Cerca de 10 desses 23 problemas foram claramente resolvidos, levando ao consenso universal, e sete foram parcialmente resolvidos, levando ao consenso parcial. Um padrão similar pode razoavelmente ser esperado na física, química e biologia. Podemos comparar esses resultados com os problemas de Os Problemas da Filosofia, de Bertrand Russell, de 1912.⁵ Nenhum destes levou à convergência universal e quase nenhum levou para qualquer coisa próxima.

É óbvio que poderia se objetar à essa tese de vários modos. Alguém poderia sustentar que tem havido mais convergência nas grandes questões filosóficas do passado do que esses estudos de caso sugerem. Alternativamente, alguém poderia sustentar que tem havido menos convergência nas grandes questões científicas do passado do que é sugerido.

Uma versão da primeira objeção nasce da observação de que disciplinas como a física eram consideradas parte da filosofia. Se voltarmos a um momento anterior à separação, então na medida em que as grandes questões da física estão entre as grandes questões da filosofia em tal momento, alta convergência na primeira levará à convergência significativa na segunda. Contudo, na medida em que a física era apenas uma parte própria da filosofia, e mais suscetível de convergência do que as outras partes, ainda se esperaria que a convergência na primeira produzisse menos convergência na segunda. E também não está inteiramente claro que a filosofia *como nós a entendemos* deveria ganhar crédito pela convergência na física: o que foi chamado de “filosofia” no passado era plausivelmente um campo diferente e mais amplo.

Em todo caso, pode-se contornar essa objeção ao se concentrar em um ponto depois da separação entre física e filosofia: digamos, 1809 ou 1909. O objetor pode responder que então haverá menos convergência somente porque nós separamos as partes da filosofia que fizeram mais progresso. Mas, isso é conceder a tese central e sustentar certa explicação dela, uma que considerarei mais tarde no artigo. Outra resposta é que haveriam mais separações depois desse

⁵ Obrigado a Jeremy Goodman por sugerir a comparação Hilbert/Russell.

ponto: psicologia, lógica, linguística e economia, por exemplo. Ainda assim, penso que as questões resolvidas por essas áreas constituem uma fração suficientemente pequena das grandes questões da filosofia em 1809 ou 1909 para que, mesmo que a filosofia ganhe crédito por elas, isso não aproxime o nível de convergência na filosofia do nível correspondente nas ciências duras.

Uma outra versão da primeira objeção sugere que algumas grandes questões filosóficas do passado alcançaram consenso e, portanto, foram retiradas da lista das grandes questões, mesmo sem terem suas próprias disciplinas se ramificando. Talvez algo assim seja plausível para algumas questões políticas e morais, por exemplo, como a questão de todas as pessoas serem iguais, onde a convergência na filosofia refletiu a convergência na sociedade de modo geral. E podem ter existido questões que foram consideradas verdadeiramente importantes em dado período (sobre a viabilidade de certas versões do idealismo, digamos) sobre as quais existe uma posição consensual hoje. A existência de questões assim ajudam contra a tese do copo vazio. Mas, no que concerne à tese do copo-meio-vazio, basta notar que a proporção de questões assim é menor na filosofia do que nas ciências duras.

Sobre a segunda objeção, poder-se-ia defender que muitas das grandes questões das ciências duras são elas mesmas questões filosóficas e têm visto pouca convergência: questões sobre a interpretação da mecânica quântica, por exemplo, ou sobre o *locus* da seleção natural. Porém, enquanto algumas das grandes questões das ciências duras não forem questões filosóficas, como certamente é plausível, e enquanto elas forem mais suscetíveis à convergência, nós ainda esperamos que a tese central seja verdadeira. Além disso, na medida em que essa objeção assume um contraste entre questões filosóficas e não-filosóficas, ela tende a reforçar o contraste subjacente na convergência.

Também poderia ser sugerido que várias questões não-filosóficas nas ciências duras encontraram baixa convergência: questões sobre a origem da vida, por exemplo. Isso certamente está correto, mas permanece plausível que um número suficiente de questões viu grande convergência, de modo que ainda existe uma diferença significativa entre o não-filosófico e o filosófico. O caso dos problemas de Hilbert na matemática destaca isso. Os problemas são na maioria não-filosóficos, e embora alguns não estejam resolvidos, a convergência geral sobre eles tem sido bastante dramática. Algo similar plausivelmente se aplica para física, química e biologia.

Uma objeção final é que ainda que exista mais acordo agora nas ciências duras do que na filosofia, essas ciências podem já começar em uma posição de maior acordo, resultando em um menor *aumento* de acordo nas ciências do que na filosofia. Essa hipótese é matematicamente consistente, mas não penso que seja especialmente plausível. Sobretudo dada a medida de acordo que é enviesada à concordância universal, como discutido antes, e dado que há discordância considerável sobre quais são as grandes questões de um período no próprio período, os vários casos em que as ciências duras (diferentemente da filosofia) se aproximarem do consenso universal tenderão a produzir igualmente maior convergência geral.

3. AS VARIEDADES DE PROGRESSO

Apesar dessa falta de convergência, é difícil negar que os *insights* de Platão e Aristóteles, Hume e Kant, Frege e Russell, Kripke e Lewis envolveram progresso filosófico significativo. Correspondentemente, minha tese do copo-meio-vazio é compatível com muitas teses diferentes de copo-meio-cheio, que asserem a existência de várias formas de progresso na filosofia. Podemos sistematizar várias dessas teses abandonando um requisito central da minha tese de cada vez.

Abandone “ampla”: tem havido (não-ampla) convergência à verdade nas grandes questões da filosofia. É plausível que tenha havido maior convergência em respostas para um pequeno número das grandes questões da filosofia: em questões sobre Deus parece ter havido maior convergência ao ateísmo, por exemplo. Também é plausível que tenha havido alguma convergência menor em respostas para muitas outras questões, como em direção ao fisicismo sobre a mente. É claro que se essa convergência contará como convergência à verdade vai depender das próprias posições filosóficas do indivíduo. Teístas e dualistas sustentarão que tais convergências constituem retrocesso em vez de progresso. Mas, se assumimos otimistamente que a convergência é de fato à verdade, pode ser que 10-20% mais filósofos tenham crenças verdadeiras sobre as respostas para as grandes questões de 1809 em 2009 do que em 1809. Se for o caso, isso é um tipo de progresso. Ainda assim, permanece plausível que a convergência seja maior em outras áreas.

Abandone “coletiva”: tem havido ampla (não-coletiva) convergência à verdade sobre as grandes questões da filosofia. A tese central é consistente com a alegação de que vários indivíduos ou sub-comunidades teriam eles mesmos amplamente convergido às respostas verdadeiras para as grandes questões. Por exemplo, em meus dias mais otimistas posso me convencer de que ao longo do tempo convergi à verdade em muitas dessas questões. Mas, sendo assim, infelizmente isso não levou à convergência coletiva à verdade. Similarmente, talvez grupos como os positivistas lógicos ou os realistas de Oxford tenham convergido à verdade. Entretanto, novamente, não sucede a convergência no nível da comunidade. Talvez tenha havido até uma ampla quantidade de convergência no nível da comunidade em certos períodos de tempo, mas, se for o caso, a convergência não persistiu.

Abandone “grande”: tem havido ampla convergência coletiva à verdade em (não-grandes) questões da filosofia. Tem havido ampla convergência em várias teses menores: a tese de que conhecimento não é crença verdadeira justificada, por exemplo, e a tese de que probabilidades condicionais não são probabilidades de condicionais. Como van Inwagen sugere na passagem acima, nós somos especialmente bons em convergir nas teses negativas que eliminam certas posições específicas. Ocasionalmente há também convergência em teses condicionais, que asserem conexões condicionais entre posições. Mas, assumo que estas não são realmente respostas para as grandes questões da filosofia.

Abandone “convergência à verdade”: têm havido amplos avanços coletivos (que não envolvem a convergência à verdade) sobre as grandes questões da filosofia. Certamente existem muitas formas de progresso filosófico que não envolvem convergência à verdade. É plausível que tenhamos hoje um entendimento bastante aumentado dos problemas implícitos nas grandes questões. Passamos a explorar novas posições e novas áreas do espaço filosófico que antes não tínhamos sequer concebido. Desenvolvemos novos métodos e argumentos melhores. Em alguns casos aplicamos a filosofia ao mundo. Certamente todas essas seriam formas de progresso. Simplesmente noto que elas não têm sido acompanhadas por ampla convergência à verdade.

Quero enfatizar que não estou apenas igualando progresso com convergência à verdade. Sou um pluralista sobre o progresso: existem muitos valores que podem ser realizados através da filosofia, e há muitas maneiras de avançar e realizar esses valores. Alcançar a verdade certamente não é o único valor. Ainda assim, certamente é um dos

valores. Disso segue que o progresso em direção à verdade é uma forma de progresso filosófico.

Mais fortemente, penso que pode ser defendido que alcançar a verdade é o objetivo primário de pelo menos muitas partes da filosofia, como a filosofia analítica. Afinal, a maior parte da filosofia, ou ao menos da filosofia analítica, consiste em propor teses como verdadeiras e argumentar pelas suas verdades. Suspeito que para a maioria dos filósofos a motivação primária para fazer filosofia é encontrar a verdade sobre os assuntos relevantes: qual é a relação entre mente e corpo? Qual é a natureza da realidade e como podemos conhecê-la? Certamente essa é a motivação primária no meu próprio caso. Portanto, sou simpático à alegação de que o progresso em direção à verdade tem certa primazia entre as formas de progresso filosófico. Mas, mesmo que se negue isso, é difícil negar que ele esteja entre essas formas.

De vez em quando é dito que uma obsessão com a verdade reflete uma concepção demasiado científica da filosofia. Nós não devemos pensar a filosofia como uma busca por respostas. Em vez disso, ela é uma busca por algo diferente: entendimento, clareza, iluminação. Concordo que esses sejam objetivos que vale perseguir, e que a filosofia pode nos ajudar a persegui-los. E posso ver porque, na ausência de respostas para os problemas filosóficos, possa parecer especialmente interessante se concentrar então em tais objetivos. No entanto, penso que devemos reconhecer que essa reação envolve algo como um rebaixamento das nossas aspirações para a filosofia. Ao menos pré-teoricamente, muitos de nós entram na filosofia procurando pela verdade e pelas respostas. Alguém poderia argumentar que essa esperança é ingênua: verdade e conhecimento não serão obtidos na filosofia, deveríamos nos contentar com algo diferente. Mas, mesmo assim, a esperança e a ingenuidade são notáveis.

Por que é a *convergência* à verdade importante, e por que deveríamos nos preocupar com sua ausência? Uma resposta óbvia é que valorizamos o conhecimento, o acordo é necessário para o conhecimento e a convergência acompanha os aumentos de conhecimento. Uma versão forte dessa posição, sugerida pela discussão de van Inwagen, é que onde há desacordo suficiente entre especialistas não se pode dizer que qualquer indivíduo sabe a verdade. Mesmo que alguns indivíduos tenham encontrado bons argumentos para conclusões verdadeiras, como eles teriam confiança justificada de que esses são bons argumentos quando tantos de seus pares discordam? Não tenho tanta certeza: penso que ao menos em alguns casos

um bom argumento pode fundamentar o conhecimento que o indivíduo tem de uma conclusão mesmo que os pares o rejeitem. Por exemplo, penso que a presença de qualquer número de pares que neguem a existência da consciência não solaparia meu conhecimento de que estou consciente. Similarmente, não solapariam argumentos que assumem essa alegação como premissa.

Porém, mesmo que o acordo não seja requerido para o conhecimento individual, algum grau de acordo plausivelmente é requerido para o conhecimento *coletivo*. Se a comunidade dos especialistas em uma questão discorda seriamente sobre a resposta da questão, então não pode ser dito que essa comunidade sabe coletivamente a resposta para a questão, e nem pode sabe-lo a comunidade mais ampla da qual ela faz parte. Mesmo quando alguns indivíduos sabem a resposta da questão, esse conhecimento individual normalmente não vai ser suficiente para o conhecimento coletivo, exceto talvez em circunstâncias especiais, tais como quando a comunidade defere a esses indivíduos.

Além disso, valorizamos o conhecimento coletivo. Uma razão pela qual o progresso das ciências duras tem sido tão impressionante é que ele plausivelmente permitiu a nós—a comunidade de investigadores—saber coletivamente as respostas dessas questões. Mas, na ausência de acordo suficiente sobre as repostas das questões filosóficas, não se pode dizer que temos conhecimento coletivo dessas respostas.

Obviamente se pode discutir quais são exatamente o grau e o padrão de acordo requeridos para o conhecimento coletivo. Contudo, é altamente plausível que o tipo de discordância que observamos sobre as respostas para as grandes questões da filosofia baste para solapar qualquer alegação de conhecimento coletivo das respostas para a maioria dessas questões. Talvez se pudesse defender que na pesquisa acima algumas posições (realismo não-cético sobre o mundo exterior, ateísmo, conhecimento a priori) exibem o tipo de consenso que permite o conhecimento coletivo. Mas, mesmo essa alegação seria ousada, e a extensão para alegações com menos consenso (fisicismo e compatibilismo, digamos) parece tão ousada que seria implausível. Portanto, assumo que a diferença na concordância sobre as grandes questões na ciência e na filosofia reflete uma diferença significativa no conhecimento coletivo que obtivemos. Semelhantemente, a diferença na convergência sobre as grandes questões reflete uma diferença significativa no aumento do conhecimento coletivo ao longo do tempo.

Isso não é negar que obtivemos uma grande quantidade de conhecimento coletivo na filosofia. Como Timothy Williamson (2006) disse, sabíamos muito mais em 2004 do que em 1964, muito mais em 1964 do que em 1924, e assim em diante. Entretanto, esse conhecimento coletivo tipicamente não envolve respostas para as grandes questões. Ele é principalmente conhecimento de respostas para questões menores, de teses negativas e condicionais, de enquadramentos disponíveis para responder essas questões, de conexões entre ideias, de modos como os argumentos sustentam ou contrariam conclusões, e por aí vai. Na ausência de convergência nas grandes questões, o conhecimento coletivo das respostas dessas questões nos escapa.

4. ARGUMENTO FILOSÓFICO

Agora me volto para a questão central: por que não há mais progresso na filosofia? E em particular: por que há menos convergência na filosofia do que nas ciências duras?

Uma explicação inicial, ainda que talvez seja meramente uma articulação do fenômeno, repousa no poder relativo dos métodos usados nesses domínios. As ciências duras têm métodos—provas, no caso da matemática, e o método observacional/experimental na física, química e biologia—que têm o poder para compelir o acordo sobre as respostas para as grandes questões. A filosofia tem um método—o método do argumento—que não o faz.

Qual é a diferença entre esses métodos? Uma diferença é que os métodos de experimento e prova partem de premissas amplamente aceitas—observações na ciência, axiomas na matemática—e prosseguem delas para conclusões fortes e surpreendentes. Aspiramos fazer isso na filosofia também: veja a observação de Russell de que o objetivo da filosofia é começar com algo tão simples que não pareça que vale a pena dizer, e terminar com algo tão paradoxal que ninguém vai acreditar (Russell 1918). Porém, na prática, premissas amplamente aceitas raramente bastam para embasar conclusões fortes e surpreendentes na filosofia.

Há certamente muitos argumentos para conclusões fortes na filosofia. Mas, na grande maioria dos casos, eles têm premissas que oponentes podem negar sem muito custo, ou inferências que oponentes podem rejeitar sem muito custo (me concentro principalmente nas

premissas, mas tudo que digo também se aplica para inferências, transformando inferências não-dedutivas em premissas tácitas de argumentos dedutivos, por exemplo). Algumas vezes a premissa negada é previamente plausível, e a negação um tanto surpreendente. Contudo, mesmo com isso a negação raramente tem a implausibilidade de se negar um axioma matemático, ou de se negar uma observação experimental bem replicada. Assim, essas negações são normalmente defensáveis, ao menos no sentido largamente sociológico de defensibilidade.

Digamos que premissas (e inferências) *consensuais* são aquelas consideradas pela comunidade como inegáveis, ou ao menos como se negá-las incorresse em custos enormes. Uma premissa consensual pode ser negada por uns poucos dissidentes, mas não pode ser objeto de discordância generalizada dentro da comunidade. Digamos que um argumento que usa apenas premissas e inferências consensuais é um argumento decisivo (note que premissas consensuais e argumentos decisivos são ambos definidos em termos sociológicos). Então a alegação é de que enquanto há argumentos decisivos para conclusões fortes na ciência, há relativamente poucos argumentos assim na filosofia.

Certamente há algumas premissas consensuais em argumentos filosóficos. Afinal, essas mesmas premissas podem incluir axiomas matemáticos e observações empíricas, bem como teoremas e teorias baseados neles. Mas, premissas matemáticas e observações empíricas sozinhas nunca são suficientes para se tirar conclusões filosóficas fortes. Premissas ou inferências adicionais são necessárias para se fazer a ponte da matemática e da ciência para a filosofia, e essas premissas e inferências são tipicamente negáveis.

Somando aos axiomas matemáticos e observações empíricas, há algumas intuições filosóficas que são extremamente difíceis de negar. Entretanto, essas intuições não são tão comuns (muitas intuições previamente plausíveis se revelam negáveis), e onde existem é tipicamente difícil tirar delas conclusões filosóficas fortes. Há alguns casos onde essas intuições, talvez em conjunção com alegações matemáticas e empíricas, nos permitem tirar conclusões fortes e surpreendentes. Isso funciona particularmente bem para teses negativas, nas quais intuições e modelos formais podem gerar contraexemplos para teses positivas ou outras razões para rejeitá-las. O argumento de Gettier partindo da intuição sobre um caso para a conclusão de que conhecimento não é crença verdadeira justificada é um exemplo. O argumento formal de Lewis de que probabilidades condicionais não são probabilidades de

condicionais talvez seja outro. Contudo, é notável que essas conclusões negativas fiquem longe de serem respostas para as grandes questões da filosofia. Quase qualquer argumento para uma resposta positiva para essas questões envolve premissas negáveis.⁶

Para a maioria dos praticantes da filosofia, o fenômeno da negabilidade das premissas é familiar em ambos os lados. Quando oferecemos argumentos para nossas posições nos frustramos em encontrar oponentes mordendo a bala⁷ ao rejeitar o que tomamos como uma premissa plausível, sem que isso sirva como qualquer sinal de derrota. Quando abordamos argumentos contra nossas posições, nós algumas vezes trabalhamos de trás pra frente, da nossa rejeição da conclusão para ver qual premissa temos que negar, e nós a negamos. No melhor dos casos, aprendemos algo com isso, e aceitamos compromissos que poderíamos previamente considerar surpreendentes. Mas, raramente esses compromissos são inviáveis para se manter.

Como resultado argumentos filosóficos tipicamente não levam ao acordo, mas ao desacordo sofisticado. Defensores de uma posição aprenderão quais compromissos extras eles precisarão aceitar para evitar os argumentos. Versões ruins de uma posição são rejeitadas e versões sofisticadas são desenvolvidas em seus lugares. Isso leva a um tipo de progresso negativo onde áreas do espaço filosófico são eliminadas, mas apenas em pequenos fragmentos de cada vez. É raro que uma posição maior geral (materialismo ou dualismo, compatibilismo ou incompatibilismo, utilitarismo ou deontologia) seja eliminada desse modo. Em vez disso,

⁶ Para recolher dados aqui, fiz uma pesquisa informal na internet com filósofos, perguntando por argumentos que sejam quase universalmente considerados por eles como estabelecendo suas conclusões. Os candidatos adicionais incluíram o argumento da doação forçosa de órgãos contra versões simples do utilitarismo, o argumento de Kripke de que a necessidade se separa da a prioricidade, o argumento de Gödel contra versões do formalismo matemático, o argumento do mal contra o teísmo, o argumento modelo-teórico contra o descritivismo global, o argumento do ator perfeito contra o behaviorismo lógico, o argumento da realizabilidade múltipla contra a teoria da identidade, o argumento de Goodman contra a lógica indutiva puramente formal, argumentos da relatividade contra o presentismo, o argumento de Frankfurt de que a responsabilidade moral não requer a capacidade de agir diferentemente, o argumento de Hart contra a teoria do comando sobre leis de Austin, a refutação de Russell da Lei Básica V de Frege, o argumento da questão aberta de Moore contra o naturalismo analítico, o argumento de Putnam para o externismo sobre o significado, o cogito de Descartes, e muitos outros. É impactante que em sua grande maioria esses argumentos sejam naturalmente considerados argumentos para conclusões negativas, dado que são argumentos contra posições bastante específicas. É óbvio que a distinção negativo/positivo não é inteiramente clara, mas temos um entendimento intuitivo razoável dela. Algumas poucas conclusões têm um sabor positivo: a própria existência (o cogito), externismo (Putnam), talvez o necessário a posteriori (Kripke) e o ateísmo (o argumento do mal). Mas, os três primeiros são no máximo casos marginais de respostas para grandes questões, e os dados da pesquisa sugerem que o segundo e o quarto são no máximo casos marginais de acordo quase universal. Tudo isso reforça o ponto de que argumentos decisivos são raros na filosofia, que argumentos decisivos para posições positivas são ainda mais raros, e que argumentos decisivos para respostas positivas para as grandes questões são tão raros que são quase inexistentes.

⁷ N. do T.: da expressão em inglês “bite the bullet”, que significa o ato de aceitar uma tese mesmo reconhecendo suas consequências implausíveis.

há grandes fragmentos sobreviventes envolvendo as posições necessárias para evitar os argumentos (materialismo tipo-B com a estratégia do conceito fenomenal, incompatibilismo de fonte, utilitarismo de dois níveis, teísmo sem benevolência ou onipotência irrestritas). O mesmo tipo de eliminação, fragmentação e refinamento frequentemente ocorre também nesses níveis inferiores. As posições que sobrevivem produzem um tipo de estrutura fractal do espaço filosófico, similar ao conjunto de Mandelbrot, com suas intrincadas complexidades em todos os níveis, mas no qual grandes regiões do espaço raramente são eliminadas inteiramente.

Esse fenômeno pode parecer um análogo filosófico da tese Duhem-Quine, em uma versão dizendo que qualquer teoria científica pode ser compatibilizada com qualquer evidência por ajustes adequados nas suposições de fundo que ligam a teoria com a evidência. Mas, na prática, teorias científicas são frequentemente rejeitadas decisivamente diante da evidência, com versões revisadas e consistentes das teorias sendo rejeitadas como indefensáveis. Com efeito, algumas inferências (não-dedutivas) da evidência para a teoria científica têm status de consenso. Teorias são eliminadas não por consenso e evidências apenas, mas por evidência consensual mais inferências consensuais. No caso filosófico, contudo, evidência consensual mais inferências consensuais são muito menos poderosos. Quando alguém argumenta contra uma teoria filosófica, normalmente há ao menos uma revisão da teoria que não é só consistente mas defensável diante da evidência consensual. Aqueles que argumentam contra uma posição filosófica às vezes acusam seus engenhosos oponentes de sustentarem um programa de pesquisa degenerado, contudo, é tipicamente muito mais difícil fazer essa objeção colar na filosofia do que na ciência. Isso pode ser porque filósofos aplicam critérios mais flexíveis para suas teorias, de maneira que as inferências que têm status consensual entre cientistas não o têm entre filósofos, mas mais plausivelmente é porque o mesmo tipo de inferência não basta para resolver questões filosóficas.

Pode ser também objetado que na ciência teorias positivas não são normalmente estabelecidas por experimentos únicos, mas por muitos experimentos coletivamente. Por paridade, podemos esperar que mesmo que posições filosóficas não sejam estabelecidas por argumentos únicos, elas possam ser estabelecidas por um número de argumentos coletivamente. Talvez existam alguns casos de teses negativas sendo estabelecidas desse modo: a rejeição de teorias da percepção dos dados dos sentidos pode ser um exemplo. Porém, mesmo esses casos são raros, e casos positivos são ainda mais raros. Na prática, se um

opponente pode rejeitar argumentos individuais para uma tese sem muito custo, ele normalmente pode rejeitar coleções de argumentos sem muito custo também.

Isso significa que todos os argumentos filosóficos para teses positivas são malsucedidos, como van Inwagen (2006) sugeriu? (van Inwagen fala de teses substanciais, mas a discussão que ele faz sugere que se requer que estas sejam teses positivas). Isso depende do que se quer dizer com “sucesso”. Se “sucesso” for definido em termos sociológicos, de maneira que o sucesso exija convencer quase todos de uma comunidade, então vimos que no melhor dos casos bem poucos argumentos filosóficos para teses positivas têm sido bem-sucedidos em nossa comunidade. Van Inwagen define “sucesso” em termos epistemológicos idealizados: um argumento bem-sucedido para uma proposição p é um que convenceria uma audiência de pensadores ideais que são inicialmente agnósticos sobre p , na presença de um oponente ideal de p . Não penso que as observações sociológicas acima (ou as observações sociológicas para as quais van Inwagen apela) chegam perto de estabelecer que nenhum argumento filosófico é bem sucedido nesse sentido (cf. McGrath e Kelly no prelo para mais sobre esse tema). Seres humanos simplesmente estão muito longe do ideal para que essa conclusão se siga.

Também não segue de nada que eu tenha dito que todos os argumentos filosóficos são petições de princípio, ou que sejam dialeticamente impotentes. Mesmo quando argumentos têm premissas negáveis, eles normalmente têm poder dialético, pois suas premissas têm um suporte prévio que não repousa sobre considerações próximas demais da conclusão. Nesse caso o argumento não é petição de princípio. Mesmo que um oponente sofisticado e comprometido vá negar a premissa, o argumento pode muito bem fazer um observador agnóstico aceitar a conclusão. Na prática, geralmente usamos esse tipo de poder dialético como critério para um bom argumento com o qual muitos filósofos podem concordar, mesmo que eles discordem sobre a persuasão definitiva do argumento.

Também não estou dizendo que esses argumentos não podem produzir conhecimento. Premissas negáveis podem ainda assim ser conhecidas por muitas pessoas como verdadeiras. Como antes, enquanto muita discordância sobre uma alegação pode solapar o conhecimento coletivo dessa alegação, isso não precisa solapar o conhecimento individual da alegação. Semelhantemente, um argumento pode fundamentar o conhecimento individual mesmo quando pares o rejeitam. Isso se aplica ainda mais quando não-pares estão envolvidos. Mesmo

com tudo que eu disse, alguns argumentos podem ter premissas e inferências que só podem ser negadas implausivelmente, ou por pensadores não-ideais. Se for assim, esses argumentos podem muito bem produzir conhecimento em seres mais razoáveis que aqueles que os negam.

Portanto, não é fácil tirar conclusões sobre a falta de força normativa a partir de premissas sobre a falta de sucesso sociológico. Talvez exista uma noção normativa intermediária, definindo um argumento bem sucedido como um que tem o poder de persuadir todos os agnósticos *competentes*, em que competência seja algum padrão razoavelmente alto mas não-ideal que muitos filósofos humanos alcançam. Há boa razão para pensar que alguns argumentos filosóficos para posições positivas persuadem todos os filósofos competentes, ou mesmo todos os agnósticos competentes. Diante disso, a discordância sobre grandes questões entre os filósofos mais capazes (por qualquer medida razoavelmente neutra) é tão predominante quanto discordâncias entre filósofos de modo geral. Isso sugere que a maioria dos argumentos filosóficos não são bem-sucedidos no sentido normativo ligado à competência, mesmo que sejam bem-sucedidos nos outros sentidos normativos.

Burton Dreben me disse uma vez de modo memorável (na única ocasião em que o encontrei, em St. Louis por volta de 1994): “Grandes filósofos não argumentam”. Ele prosseguiu elaborando que entre Frege, Russell, Wittgenstein, Carnap ou Quine, ninguém realmente oferece argumentos para suas posições. Claro que isso não é estritamente verdadeiro, contudo, penso que o ponto dele era que nesses filósofos o trabalho real não era feito por argumentos para uma tese, mas pela própria tese, ou o enquadramento em que ela estava encaixada. Uma versão refinada dessa alegação (sugerida para mim por Gene Callahan) pode dizer: grandes filósofos podem argumentar, mas seus argumentos não são o que os torna grandes. Uma parte do pensamento de Dreben, como eu o entendi, era de que como argumentos são facilmente refutados, dar argumentos é um sinal de fraqueza. É melhor simplesmente asserir e desenvolver uma tese. Assim os leitores têm que se envolver com a própria tese, sem a distração barata de refutar argumentos para ela.

Rawls (2001) elabora a posição de Dreben em uma direção algo diferente: “Burt não negaria, é claro, o fato evidente de que filósofos fazem muitos argumentos complicados. Mas ele pensa que no fundo não há argumentos que um filósofo possa usar para convencer outro de um ponto metafísico. No nível básico, filósofos simplesmente se apoiam em e apelam para ‘dados’ diferentes. É uma disputa sem resolução por argumentos. Burt disse que Quine é um

metafísico, um metafísico da ciência. Com isso ele quis dizer que Quine não argumenta pelo fisicismo, ou realismo científico. Ele o assume e elabora sua posição a partir daí.”

Descobri que para mim mesmo é impossível seguir o conselho de Dreben. Em meu trabalho sou um argumentador compulsivo, o que sem dúvida me torna alvo para um *modus tollens* da tese de Dreben. Mas, certamente é raro que esses argumentos convençam um grande setor da população. Isso é especialmente o caso quando muitas das pessoas já têm compromissos firmes, em questões como o problema mente-corpo e a teoria do significado: aí é difícil fazer mais que convencer algumas pessoas aqui e ali. Sobre questões em que as pessoas são inicialmente agnósticas ou seus compromissos são fracos, pode haver mais movimento.⁸ Mesmo aqui o fato de que o movimento é limitado reforça o ponto básico.

A lição é que o consenso na filosofia está tão difícil de se obter quanto sempre foi, e argumentos decisivos são tão raros quanto sempre foram. Para mim, esse é o maior desapontamento na prática da filosofia. Uma vez que se fez filosofia por algum tempo, não se espera mais que argumentos produzam acordo, e se julga um argumento bom quando ele meramente tem algum poder dialético. No entanto, esse é um ajuste de expectativas em resposta à uma realidade desapontadora. Antes de fazer filosofia se poderia esperar que algo mais era possível.

5. NOVOS MÉTODOS FILOSÓFICOS

Diante da falha dos métodos filosóficos tradicionais, podemos procurar por novos métodos. Ocasionalmente novos métodos são desenvolvidos por pensadores que se consideravam filósofos e que ajudaram a resolver questões que foram uma vez consideradas filosóficas: veja o desenvolvimento da lógica, da física, da psicologia, e por aí vai. É natural esperar que novos métodos possam produzir mais progresso.

⁸ Para praticar autobiografia: tenho a impressão de que meus argumentos com Andy Clark para a tese da mente estendida (uma área em que compromissos prévios eram relativamente fracos) pode ter convencido mais gente que meus argumentos contra o fisicismo ou para a semântica bidimensional (áreas em que compromissos prévios são fortes). Mesmo aqui suspeito que a tese e o enquadramento convenceram tantas pessoas quanto os argumentos. Talvez o argumento mais efetivo de todos tenha sido o de “The Matrix as Metaphysics”, que convence muita gente da posição de que se estamos em um cenário de matrix ou em que somos cérebros em cubas, a maioria das nossas crenças são verdadeiras (ao menos ele tem esse efeito em apresentações de aula e discussões informais; houve relativamente pouca discussão impressa do argumento). Ainda que as pessoas achem essa posição inicialmente contra-intuitiva, se revela que o compromisso prévio delas era fraco.

Por volta do último século, muitos métodos filosóficos novos foram desenvolvidos e muitos métodos filosóficos velhos foram refinados para ajudar a alcançar conclusões filosóficas. A filosofia empírica se apoia na ciência empírica. A filosofia formal se apoia no raciocínio formal. A filosofia linguística se apoia na análise da linguagem. A fenomenologia se apoia na reflexão fenomenológica. A filosofia feminista se apoia na consideração e análise de gêneros. A filosofia transcultural se apoia em múltiplas tradições culturais na filosofia. A filosofia experimental se apoia no estudo empírico dos juízos filosóficos.⁹

Todos esses métodos levaram a novos *insights* e ao progresso filosófico. Todos levaram a novos argumentos e conclusões interessantes. Porém, manifestamente, nenhum desses métodos levou à qualquer convergência recente sobre as respostas das grandes questões da filosofia. No surgimento de cada um desses métodos a discordância filosófica era tão predominante quanto sempre foi. Mesmo dentro de uma tradição são poucos os casos em que as grandes questões são tidas como resolvidas. Em vez disso, esses métodos nos levaram novamente para versões mais sofisticadas de discordâncias antigas.

Em muitos casos o problema básico é o da negação das premissas. Nos casos da filosofia empírica e formal, já vimos que resultados empíricos e formais precisam ser combinados com premissas de ligação adicionais para resolver uma questão filosófica. Na maioria dos casos vemos que essas premissas de ligação podem ser negadas. Frequentemente elas são tão controversas quanto as conclusões que buscam estabelecer. Em alguns casos resultados empíricos e formais ajudam a resolver questões relativamente pequenas, bem como introduzem e abordam questões novas importantes. Mas, quando elas são trazidas para considerações sobre as grandes questões, é raro que façam muito para produzir consenso.

Há algumas exceções parciais: talvez o peso da evolução sobre o teísmo, da relatividade sobre o presentismo e do teorema de Gödel no formalismo matemático. Mas, essas exceções não são especialmente comuns, e mesmo nesses casos há versões modificadas das posições relevantes que retiveram vários defensores sérios. As exceções que existem parecem essencialmente cair em duas classes. Primeiro, há casos em que métodos empíricos pesam fortemente sobre áreas da filosofia que se concentram na realidade concreta, como subcampos da metafísica e da filosofia da ciência. Segundo, há casos em que métodos formais

⁹ E há muitos outros métodos que não tenho competência para discutir. Por exemplo, Nielsen (1987) sugere que uma esperança de progresso na filosofia é a teoria crítica.

pesam fortemente em áreas que lidam com questões formais, como a filosofia da matemática e da lógica. Óbvio que ambos os tipos de método são frequentemente trazidos para serem considerados em outras áreas—áreas normativas como ética e epistemologia, por exemplo—mas casos em que produzem consenso são muito mais raros.

Se pode pensar que a filosofia da mente seria uma exceção, dadas as óbvias conexões com a neurociência e a psicologia. Mas, mesmo nela, essas ciências parecem ter deixado as grandes questões—os problemas da consciência e da intencionalidade, da causação mental e do livre arbítrio—bastante abertas. Certamente tem havido argumentos da neurociência e da psicologia para posições sobre esses problemas, mas na maioria dos casos as premissas de ligação requeridas têm sido tão controversas quanto a maioria das outras alegações filosóficas. O que resultou foi uma sofisticação muito aumentada com versões cientificamente informadas de posições relevantes, mas não muito mais em questão de consenso. Talvez o maior peso sobre essas grandes questões acerca da mente não tenha vindo desses campos, mas da física, onde a evidência do fechamento causal no nível microfísico colocou uma séria pressão sobre posições como o dualismo interacionista. Entretanto, mesmo nisso muitos têm resistido à pressão, e em qualquer caso a negação do interacionismo não leva realmente à uma posição positiva.

Outros métodos novos sequer oferecem premissas com a relativa segurança das premissas empíricas e formais. Em fenomenologia, por exemplo, as premissas fenomenológicas centrais são tipicamente tão negáveis quanto qualquer outra premissa filosófica. Algo similar pode se aplicar para a filosofia feminista e transcultural, enquanto outros métodos podem exibir uma mistura dos dois padrões acima.¹⁰ Alguns métodos, como a filosofia feminista e a filosofia experimental, têm cumprido um papel crítico crucial, mas a lição tem sido amplamente a de diminuirmos nossa confiança sobre as respostas das grandes questões em vez de fortalecê-la.

Claro, novos métodos estão sempre sendo desenvolvidos. Ainda acontece de problemas gradualmente migrarem da filosofia para a ciência conforme os métodos se tornam mais rigorosos e decisivos: dois exemplos recentes incluem o desenvolvimento da semântica formal e o desenvolvimento em andamento de uma ciência da consciência. Ainda assim,

¹⁰ Sobre minha própria opinião sobre os poderes e limites da filosofia experimental e linguística, veja <http://consc.net/papers/xphi.pdf> e <http://consc.net/papers/langphil.pdf> respectivamente.

mesmo nesses casos seria difícil dizer que os novos métodos levaram ao consenso sobre as grandes questões filosóficas que precediam seus desenvolvimentos. Portanto, enquanto podemos esperar por mais métodos que produzam convergência sobre as grandes questões, esses métodos terão que ir bem além do que vimos no último século.

6. EXPLICAÇÕES

Até agora dei uma explicação muito parcial da relativa falta de convergência na filosofia. Há menos convergência na filosofia porque o método filosófico tem menos poder para compelir ao acordo, e tem menos poder por causa do fenômeno da negabilidade das premissas: argumentos para conclusões fortes na filosofia (diferentemente da ciência e da matemática) quase sempre têm premissas ou inferências que podem ser rejeitadas sem muito custo.

Contudo, essa explicação fica bem perto da superfície do fenômeno. É natural pedir uma explicação mais profunda. Por que os argumentos de premissas consensuais são relativamente incapazes de resolver as grandes questões da filosofia? E de modo mais geral, por que há tão pouca convergência na filosofia?

(1) *Especiação disciplinar*. A resposta mais popular para essa pergunta, ao menos entre filósofos, é que o campo está sujeito à um tipo de especiação disciplinar. Como já discuti, muitas disciplinas novas nasceram da filosofia ao longo dos anos: física, psicologia, lógica, linguística, economia, e por aí vai. Em cada caso, esses campos nasceram como ferramentas que foram desenvolvidas para abordar questões de forma mais precisa e decisiva. A tese chave é que quando desenvolvemos métodos para resolver conclusivamente questões filosóficas, esses métodos constituem um novo campo e as questões não são mais consideradas filosóficas. Logo, é de se esperar que as questões que restam estejam menos sujeitas ao acordo do que as que estão em outras disciplinas.

Certamente há algum valor nessa explicação. A tese chave é plausível e a tese central parece seguir dela. Ainda assim, penso que há alguns limites nessa explicação.

Já notei um limite: os campos que se separaram nem sempre responderam as grandes questões filosóficas que os precederam. A psicologia não fez muito para resolver o problema mente-corpo, por exemplo, e a linguística não resolveu realmente as questões filosóficas mais profundas sobre o significado. Lógica e física chegaram mais perto, mas mesmo aí é defensável que elas não resolveram algumas das maiores questões filosóficas precedentes. Agora, pode ser dito que as questões filosóficas menores que esses campos resolvem ainda assim correspondem às grandes questões nos novos campos, assim explicando a tese central sobre a convergência relativa. Porém, desejamos uma explicação de porque a questão filosófica *precedente* na filosofia tem sido tão difícil de responder. Na medida em que essas grandes questões não têm sido resolvidas pela especialização disciplinar a especialização não pode responder essa questão.

A objeção mais geral é que ainda que a tese da especialização possa explicar, *de dicto*, porque há menor convergência nas grandes questões da filosofia do que nas grandes questões de outras áreas, nós também queremos uma explicação *de re*, envolvendo essas grandes questões, sobre porque *elas* em particular recebem tão pouca convergência. Uma hipótese é que isso é simplesmente uma questão de sorte: todas as questões são igualmente aptas à convergência, e através de sorte aleatória algumas receberam convergência (e com isso se especializaram) antes que outras. Mas, deixando de lado essa hipótese implausível, a relativa falta de convergência sobre essas questões é presumivelmente explicada por algo distinto sobre essas questões e a relação delas conosco. Podemos então perguntar qual é exatamente essa característica ou características dessas questões que explica(m) a falta de convergência. A especialização não tem poder para responder essa pergunta, portanto outra resposta é necessária.

(2) *Antirrealismo*. Uma resposta é que não há convergência à verdade porque não há verdades objetivas a serem encontradas nos domínios relevantes. Onde há verdade objetiva, ela serve como um tipo de ímã para a convergência, mas em sua ausência há apenas um corpo desregrado de opiniões que não devemos esperar que convirja. Muitos filósofos terão simpatia por essa linha em algumas áreas. Tenho simpatia pelo antirrealismo sobre a ética e sobre algumas questões em ontologia. Ainda assim, isso deixa muito para se ser realista sobre. E mesmo aceitar o antirrealismo moral, digamos, deixa em aberto o porquê de haver tão pouca convergência na própria questão do realismo moral. É claro que permanece a possibilidade do

antirrealismo filosófico global, mas essa não é uma posição especialmente plausível ou atraente.

- (3) *Disputas verbais*. Outra resposta é a de que há pouca convergência porque os participantes estão em um diálogo de surdos. Cada lado está usando termos-chave de modos diferentes e cada um está correto no que seu próprio uso diz respeito. Em “Verbal Disputes” argumentei que disputas verbais são comuns na filosofia. Por exemplo, penso que muitos debates na filosofia do livre arbítrio e na filosofia da linguagem têm um elemento verbal significativo. E penso que resolver disputas verbais pode levar ao progresso filosófico. Contudo, frequentemente esclarecemos os termos-chave em uma disputa parcialmente verbal, e descobrimos que resta uma disputa substancial. E há um núcleo de questões fundamentais (incluindo muitas questões normativas, bem como o problema mente-corpo e outros problemas envolvendo conceitos filosóficos “bedrock”, nos termos de “Verbal Disputes”) para os quais o diagnóstico de uma disputa verbal parece bastante implausível.
- (4) *Maior distância dos dados*. Uma resposta naturalmente sugerida pela discussão dos argumentos decisivos é que há menos convergência na filosofia do que na ciência porque a filosofia tende a se ocupar de domínios distantes de dados claros. Colocando isso de modo quineano, teses filosóficas estão longe da periferia da rede de crenças. Ainda assim, aparentemente o mesmo também vale para muitas alegações altamente teóricas na ciência, por exemplo, envolvendo o passado longínquo e o que é muito pequeno. E plausivelmente o mesmo vale para a matemática. Nesse caso se poderia apontar para axiomas matemáticos e intuições como dados, porém, isso então levanta a questão de por que nós não temos dados filosóficos análogos para resolver questões filosóficas. Logo, essa opção tende a reformular o problema em vez de solucioná-lo.¹¹
- (5) *Explicações sociológicas*. É natural supor que fatores sociológicos desempenhem um papel no impedimento da convergência. Quando nossos argumentos não são universalmente aceitos, frequentemente creditamos isso ao histórico profissional dos nossos oponentes, ou aos pressupostos falsos que estão disseminados na

¹¹ Também cabe notar (como Larry Solum me sugeriu) que as ciências sociais têm muito menos convergência que as ciências duras apesar de estarem menos distantes dos dados que a filosofia. Uma questão geral interessante é se a falta de convergência nas ciências sociais e na filosofia deveria receber explicações diferentes ou uma explicação uniforme. Suspeito que seja a primeira opção: por exemplo, a complexidade e desordem dos sistemas sociais parece especialmente relevante nas ciências sociais, mas menos relevante na filosofia.

profissão, ou ao vínculo profissional ou emocional às posições alternativas. Penso que não há como negar que fatores sociológicos desempenham um papel importante na determinação das posições filosóficas amplamente aceitas em um período. A impopularidade da distinção analítico-sintético nas décadas após “Dois Dogmas do Empirismo” certamente não foi decorrelacionada com a posição de poder de Quine na profissão e com seu impacto nos estudantes de graduação. Ainda assim, muitos dos fatores sociológicos relevantes também agem nas ciências. Portanto, para explicar a diferença com as ciências se teria que apontar para diferenças sociológicas relevantes, ou se combinar a explicação sociológica com outras características distintas das questões filosóficas. Algumas diferenças sociológicas potenciais incluem a hipótese de que filósofos são mais recompensados por discordâncias do que cientistas, que são mais tolerantes com a dissidência, ou que foram treinados para terem padrões mais altos para a aceitação de posições. Se poderia também apontar para diferenças de financiamento, treinamento e estruturas de pesquisa. Ainda assim, é difícil crer que a diferença em convergência entre, digamos, o projeto do genoma humano e o problema mente-corpo meramente se resume a esses fatores sociológicos. Portanto, explicações sociológicas funcionam melhor quando são combinadas com teses adicionais sobre a peculiaridade das questões filosóficas. Por exemplo, se poderia sugerir que a maior distância entre os dados e as teses filosóficas facilita para que a resistência sociologicamente fundamentada se sustente.

- (6) *Explicações psicológicas.* As explicações psicológicas estão intimamente relacionadas com as sociológicas, sustentando que há algo distinto sobre mentes humanas ou mentes de filósofos que impede a convergência em questões filosóficas. Talvez exista algum defeito psicológico que nos impede de reconhecer verdades filosóficas, por exemplo. Em algum nível alguma explicação como essa tem que ser parte da história: se ao menos pensadores ideais poderiam convergir à verdade, então nossa falha pode ser marcada pela não-idealidade dos nossos raciocínios. Mas, aí a questão crucial será: quais são os aspectos em que nosso raciocínio é não-ideal, e quais são os aspectos em que as questões filosóficas são peculiares, tais que a não-idealidade dos nossos raciocínios nos impede de convergir à verdade em questões filosóficas?

(7) *Explicações evolutivas*. Às vezes é sugerido (e.g. McGinn, 1993) que há uma explicação darwiniana para a falta de progresso na filosofia. A ideia básica é que não evoluímos para sermos bons em filosofia, dado que no ambiente evolutivo não houve pressão seletiva que favorecesse a habilidade filosófica ou qualquer coisa que se correlacionasse fortemente com ela. Talvez haja alguma plausibilidade nisso, ainda que seja necessário algum trabalho para se explicar porque o mesmo não se aplica para habilidade de praticar matemática abstrata ou ciência altamente teórica. Em qualquer caso, esse tipo de explicação vai funcionar melhor em conjunção com a explicação psicológica, e levanta as mesmas questões cruciais discutidas sobre ela.

Penso que todas essas sete explicações podem estar parcialmente corretas. Contudo, não penso que do modo como estão elas coletivamente forneçam uma explicação completa do fenômeno. Para fazer isso muitos dos detalhes precisariam ser preenchidos. Em particular, ainda precisaríamos de uma boa explicação do que exatamente é peculiar sobre questões filosóficas de modo que elas levem à falta de convergência. Apenas (2) e (4) realmente abordam isso, mas (antirrealismo global à parte) (2) se aplica apenas em alguns casos, enquanto (4) está perto demais de ser uma reformulação do fenômeno. Deve ser possível fornecer uma explicação dessa peculiaridade que se misture com as explicações psicológicas, evolutivas e talvez sociológicas para fornecer uma explicação completa da falta de convergência. Porém, por ora penso que isso continua uma questão aberta.

7. OS PROSPECTOS PARA MAIS PROGRESSO

Finalmente: quais são os prospectos para mais progresso filosófico? É possível que possamos eventualmente convergir à verdade nas grandes questões da filosofia?

Para discutirmos isso precisamos abordar a questão das respostas dessas questões serem mesmo cognoscíveis em princípio, por agentes cognitivos suficientemente ideais. Aqui vou apenas indicar minha própria posição positiva sobre essa questão. Em *Constructing the World* argumentei por uma tese da scrutabilidade (chamada “Escrutabilidade Fundamental” no livro), mantendo que todas as verdades são implicadas a priori por verdades empíricas

fundamentais acerca de propriedades e leis fundamentais e naturais. Segue (simplicemente) que se alguém pudesse saber todas as verdades empíricas fundamentais e raciocinar idealmente, esse alguém poderia conhecer todas as verdades, incluindo todas as verdades filosóficas.

É claro que a tese da scrutabilidade pode ser negada. Se ela for falsa, então mesmo raciocínios ideais partindo de verdades empíricas fundamentais podem não nos permitir conhecer a verdade filosófica. Se poderia preservar uma versão modificada da tese ao se expandir as verdades fundamentais na base para incluir certas verdades filosóficas: princípios ontológicos e normativos fundamentais, digamos. Mas então as próprias verdades fundamentais poderiam ficar além de um alcance epistemológico ideal. De qualquer jeito, se verdades filosóficas não são scrutáveis de uma base apropriada, não devemos esperar convergência à verdade mesmo em uma comunidade de agentes cognitivos ideais.

Se a tese da scrutabilidade for verdadeira, por outro lado, uma posição mais positiva ganha força. A tese não implica que podemos conhecer todas as verdades filosóficas, mas fornece uma maneira útil de classificar os casos em que ficamos longe de conhecer, e de classificar, mais geralmente, casos em que falhamos em convergir. Primeiro, há casos de antirrealismo sobre um domínio, onde não há verdade filosófica a ser conhecida. Segundo, há casos onde múltiplos grupos sabem as verdades filosóficas, mas em que as disputas verbais os impedem de reconhecer seus acordos. Terceiro, há casos em que somos ignorantes sobre as verdades empíricas fundamentais. Quarto, há casos em que nosso raciocínio é não-ideal.

Penso que muitos dos casos mais difíceis na filosofia caem na última categoria: questões cujas respostas são cognoscíveis para agentes cognitivos ideais, mas (ainda) não para nós. Isso então levanta a pergunta chave: as respostas são cognoscíveis ou incognoscíveis para humanos?

McGinn (1993) e van Inwagen (2009) advogaram a incognoscibilidade: humanos simplesmente não são inteligentes o bastante para responder as grandes questões. A ideia é que há algum nível de inteligência ou aptidão que seria suficiente para responder essas questões, mas os humanos ficam abaixo desse nível.

Van Inwagen argumenta para essa conclusão como segue. Ele sugere que é implausível que estejamos muito acima desse nível, dada a falta de progresso até o momento,

e que é antecedentemente improvável que estejamos praticamente nesse nível. Então, é muito mais provável que o nível fique acima de nós. Não estou tão certo sobre esse argumento. Penso que já sabemos que para um grande número de questões humanas estão exatamente no nível para ir bem: questões científicas e matemáticas, por exemplo. Por causa disso, é defensável que estejamos em um limiar especial de inteligência no qual um número extraordinariamente amplo de questões fica ao nosso alcance ao longo do tempo. Não é óbvio se questões filosóficas irão ou não cair dentro desse alcance, mas não é obviamente mais provável que elas não caiam do que caiam.

Se McGinn e van Inwagen estão certos, permanece possível que possamos responder questões filosóficas ao aprimorar primeiramente nosso nível de inteligência, talvez por melhoria ou extensão cognitiva. Alternativamente, poderíamos construir seres com inteligência artificial mais inteligentes que nós, que seriam então capazes de construir seres com inteligência artificial mais inteligentes que eles, e assim em diante. A explosão de inteligência resultante poderia levar até criaturas que poderiam finalmente responder as grandes questões filosóficas.

Se McGinn e van Inwagen estão errados, por outro lado, então podemos eventualmente responder questões filosóficas sem melhorias cognitivas radicais. Talvez tenhamos que desenvolver novos métodos, aumentar a disciplina, ter novos tipos de *insights*, e talvez vá precisar haver uma revolução conceitual ou duas, mas nada disso vai ficar fora da capacidade humana. Pode se revelar que há uma curva de sofisticação filosófica crescente tal que, depois de certo ponto na curva, um progresso maior é possível. Não estamos lá ainda, mas estamos trabalhando na direção disso.

Não é óbvio se McGinn ou van Inwagen estão certos ou errados. A questão das grandes questões filosóficas serem humanamente solúveis é ela mesma uma grande questão metafilosófica. Como outras grandes questões da filosofia, é uma para a qual nós atualmente não sabemos a resposta. Ambas as respostas para essa questão metafilosófica parecem estar abertas, e nós atualmente não temos fortes razões para preferir nenhuma.

Se nós não sabemos qual dessas duas opções é o caso, então penso que para fazer filosofia nós podemos assumir a suposição de trabalho de que é a segunda opção: as questões podem ser respondidas por nós, mas até o momento não estão resolvidas. Então nós podemos simplesmente fazer filosofia tão bem quanto podemos, fazendo nosso melhor para pensar

nesses novos *insights*, métodos e conceitos que possam finalmente nos levar a responder as questões. Afinal, ainda estamos aprendendo a fazer filosofia bem. Para ver o quão longe ela pode nos levar, temos que continuar fazendo filosofia.

BIBLIOGRAFIA

BOURGET, D. e CHALMERS, D. J. 2014. What do philosophers believe? *Philosophical Studies*.

CHALMERS, D. J. 2011. Verbal Disputes. *Philosophical Review* 120: 515-66.

CHALMERS, D. J. 2012. *Constructing the World*. Oxford University Press.

DIETRICH, E. 2010. There is no progress in philosophy. *Essays in Philosophy* 12: 329-44.

HILBERT, D. 1902. Mathematical Problems. *Bulletin of the American Mathematical Society* 8: 437-79.

KELLY, T. e MCGRATH, S. no prelo. Are there any successful philosophical arguments? In (J. Keller ed.) *Being, Freedom, and Method: Themes from van Inwagen*. Oxford University Press.

KRIPPENDORFF, K. 2013. *Content Analysis: An Introduction to its Methodology*, 3rd edition. Sage Publishing.

MCGINN, C. 1993. *Problems in Philosophy*. Oxford University Press.

MOODY, T. 1986. Progress in philosophy. *American Philosophical Quarterly* 23: 35-46.

NIELSEN, K. 1987. Can there be progress in philosophy? *Metaphilosophy* 18: 1-30.

RAWLS, J. 2001. Afterword. In (J. Floyd & S. Shieh eds.) *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy*. Oxford University Press.

RUSSELL, B. 1912. *The Problems of Philosophy*. Williams and Norgate.

RUSSELL, B. 1918. "The Philosophy of Logical Atomism". *The Monist* 28: 295-527. Reimpresso como *The Philosophy of Logical Atomism* (Taylor and Francis, 2009).

STOLJAR, D. no prelo. *Philosophical Progress: In Defense of a Reasonable Optimism*.

VAN INWAGEN, P. 2004. Freedom to break the laws. *Midwest Studies in Philosophy* 28: 334-50.

VAN INWAGEN, P. 2006. *The Problem of Evil: The Gifford Lectures*. Oxford University Press.

VAN INWAGEN, P. 2009. *Metaphysics* (third edition).

WILLIAMSON, T. 2006. Must do better. In (P. Greenough & M. Lynch, eds.) *Truth and Realism*. Oxford University Press.

YANDELL, B. H. 2002. *The Honors Class: Hilbert's Problems and their Solvers*. A. K. Peters.

LÓGICA MODAL*

James Garson

Tradução de Heitor Hedler Siqueira Campos¹

Um modal é uma expressão (como 'necessariamente' ou 'possivelmente') que é usada para qualificar a verdade de um juízo. A lógica modal é, estritamente falando, o estudo do comportamento dedutivo das expressões 'é necessário que' e 'é possível que'. Entretanto, o termo 'lógica modal' pode ser usado de forma mais abrangente para uma família de sistemas similares. Estes incluem lógicas doxásticas, lógicas para expressões temporais, lógicas para expressões deônticas (morais) tais como 'é obrigatório que' e 'é permitido que', e muitas outras. O entendimento das lógicas modais é particularmente valioso na análise formal do argumento filosófico, onde expressões da família modal são tanto comuns quanto ambíguas. A lógica modal também possui importantes aplicações na ciência da computação.

1 O que é a Lógica Modal?

Interpretada de forma estrita, a lógica modal estuda o raciocínio que envolve o uso das expressões 'necessariamente' e 'possivelmente'. Entretanto, o termo 'lógica modal' é usado de forma mais abrangente para cobrir uma família de lógicas que possuem regras similares e uma diversidade de símbolos diferentes.

Segue uma lista descrevendo as lógicas mais conhecidas desta família.

Lógica	Símbolos	Expressões Simbolizadas
Lógica Modal	□	É necessário que...
	◇	É possível que...

¹ Mestrando da UFPE. Revisor 1: Márcio Moretto Ribeiro, professor doutor da EACH-USP. Revisor 2: João Marcos, professor adjunto, UFRN.

Lógica Deôntica	O	É obrigatório que...
	P	É permitido que....
	F	É proibido que...
Lógica Temporal	G	Será sempre o caso que...
	F	Será o caso que...
	H	Sempre foi o caso que...
	P	Foi o caso que...
Lógica Doxástica	Bx	x acredita que...

2 Lógicas Modais

As lógicas mais conhecidas na família modal são construídas a partir de uma lógica mais fraca **K** (em homenagem a Saul Kripke). Na leitura mais estrita, a lógica modal se preocupa com a possibilidade e a necessidade. Uma diversidade de símbolos diferentes pode ser desenvolvida para tais lógicas utilizando **K** como fundamento. Os símbolos de **K** incluem ' \sim ' para 'não', ' \rightarrow ' para 'se...então', e ' \Box ' para o operador modal 'é necessário que'. (Os conectivos '&', ' \vee ' e ' \leftrightarrow ' podem ser definidos a partir de ' \sim ' e ' \rightarrow ' como na lógica proposicional.) **K** resulta da adição aos princípios da lógica proposicional o seguinte.

Regra da Necessitação: Se A é um teorema de **K**, então $\Box A$ também o é.

Axioma de Distribuição: $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

(Nesses princípios utilizamos ' A ' e ' B ' como metavariáveis sobre fórmulas da linguagem.) De acordo com a Regra da Necessitação, qualquer teorema da lógica é necessário. O Axioma da Distribuição diz que se é necessário que se A então B , então se necessariamente A então necessariamente B .

O operador \Diamond (para 'possivelmente') pode ser definido a partir de \Box tomando $\Diamond A = \sim \Box \sim A$. Em **K**, os operadores \Box e \Diamond comportam-se de forma muito semelhante aos

quantificadores \forall (todo) e \exists (algum). Por exemplo, a definição de \diamond a partir de \square reflete a equivalência de $\forall xA$ com $\sim\exists x\sim A$ na lógica de predicados. Por conseguinte, $\square(A\&B)$ implica $\square A\&\square B$ e vice-versa; enquanto $\square A\vee\square B$ implica $\square(A\vee B)$, mas *não* vice-versa. Isso reflete os padrões exibidos pelo quantificador universal: $\forall x(A\&B)$ implica $\forall xA\&\forall xB$ e vice-versa, enquanto $\forall xA\vee\forall xB$ implica $\forall x(A\vee B)$ mas não vice-versa. Paralelos similares entre \diamond e \exists podem ser traçados. A base para essa correspondência entre os operadores modais e os quantificadores irá emergir mais claramente na seção sobre Semântica de Mundos Possíveis.

O sistema **K** é muito fraco para prover uma explicação adequada da necessidade. O seguinte axioma não é demonstrável em **K**, mas é claramente desejável.

$$(M) \square A \rightarrow A$$

(M) enuncia que o que quer que seja necessário é o caso. Note que (M) seria incorreto se \square fosse lido 'deve ser o caso que' ou 'era o caso que'. Então a presença do axioma (M) distingue a lógica modal correspondente de outras lógicas na família modal. Uma lógica modal *M* resulta da adição de (M) a **K**. (Alguns autores chamam esse sistema de **T**.)

Muitos lógicos acreditam que *M* ainda é fraca demais para formalizar corretamente a lógica da necessidade e da possibilidade. Eles recomendam ulteriores axiomas para governar a iteração, ou repetição, dos operadores modais. Seguem os dois mais famosos axiomas de iteração:

$$(4) \square A \rightarrow \square \square A$$

$$(5) \diamond A \rightarrow \square \diamond A$$

S4 é o sistema que resulta de adicionar (4) a *M*. De forma similar, **S5** é *M* adicionado de (5). Em **S4**, a sentença $\square \square A$ é equivalente a $\square A$. Como resultado, qualquer cadeia de quadrados pode ser substituída por um único quadrado, e o mesmo ocorre para cadeias de losangos. Isso equivale à ideia de que a iteração de operadores modais é supérflua. Dizer que *A* é necessariamente necessário é considerada uma maneira inutilmente longa de dizer que *A* é necessário. O sistema **S5** possui princípios ainda mais fortes para simplificar cadeias de operadores modais. Em **S4**, uma cadeia de operadores *do mesmo tipo* pode ser substituída por aquele operador; em **S5**, cadeias contendo tanto quadrados quanto losangos são equivalentes

ao último operador na cadeia. Assim, por exemplo, dizer que é possível que A seja necessário é o mesmo que dizer que A é necessário. Segue um sumário dessas características de **S4** e **S5**.

$$\mathbf{S4}: \Box\Box\dots\Box = \Box \text{ e } \Diamond\Diamond\dots\Diamond = \Diamond$$

$$\mathbf{S5}: 00\dots\Box = \Box \text{ e } 00\dots\Diamond = \Diamond, \text{ onde cada } 0 \text{ é um } \Box \text{ ou um } \Diamond$$

Pode-se se engajar em infundáveis discussões sobre a correção ou incorreção desses e outros princípios de iteração para \Box e \Diamond . A controvérsia pode ser parcialmente resolvida pelo reconhecimento de que as palavras 'necessariamente' e 'possivelmente' possuem muitos usos diferentes. Então, a aceitabilidade dos axiomas da lógica modal depende de quais desses usos temos em mente. Por esse motivo, não há uma lógica modal, mas ao invés disso uma família inteira de sistemas construídos a partir de M . O relacionamento entre esses sistemas é diagramado na Seção 8, e sua aplicação a diferentes usos de 'necessariamente' e 'possivelmente' podem ser mais profundamente entendida estudando sua semântica de mundos possíveis na Seção 6.

O sistema **B** (em homenagem ao lógico Brouwer) é formado adicionando o axioma (B) a M .

$$(B) A \rightarrow \Box\Diamond A$$

É interessante notar que **S5** pode ser formulado de forma equivalente adicionando (B) a **S4**. O axioma (B) levanta um ponto importante sobre a interpretação das fórmulas modais. (B) diz que se A é o caso, então A é necessariamente possível. Pode-se sustentar que (B) deveria sempre ser adotado em qualquer lógica modal, pois certamente se A é o caso, então é necessário que A seja possível. Entretanto, há um problema com essa assertiva que pode ser exposto notando que $\Diamond\Box A \rightarrow A$ é demonstrável a partir de (B). Então $\Diamond\Box A \rightarrow A$ deveria ser aceitável se (B) o for. Todavia, $\Diamond\Box A \rightarrow A$ diz que se A é possivelmente necessário, então A é o caso, e isso está longe de ser óbvio. Por que (B) parece óbvio, enquanto uma de suas consequências parece não ser realmente óbvia? A resposta é que há uma ambiguidade perigosa na interpretação em português de $A \rightarrow \Box\Diamond A$. Frequentemente utilizamos a expressão 'Se A então necessariamente B ' para expressar que o condicional 'se A então B ' é necessário. Essa interpretação corresponde a $\Box(A \rightarrow B)$. Em outras ocasiões, queremos dizer que se A , então B é necessário: $A \rightarrow \Box B$. Em português, 'necessariamente' é um advérbio, e já que

advérbios são usualmente empregados próximos a verbos, não possuímos nenhuma maneira natural de indicar se o operador modal se aplica a todo o condicional, ou a seu conseqüente. Por tais motivos, há uma tendência de confundir $(B): A \rightarrow \Box \Diamond A$ com $\Box(A \rightarrow \Diamond A)$. Mas $\Box(A \rightarrow \Diamond A)$ não é o mesmo que (B) , pois $\Box(A \rightarrow \Diamond A)$ já é um teorema de M , enquanto (B) não o é. É preciso tomar um cuidado especial para que nossa reação positiva a $\Box(A \rightarrow \Diamond A)$ não afete nossa avaliação de (B) . Uma maneira simples de nos protegermos é formular B de maneira equivalente utilizando o axioma $\Diamond \Box A \rightarrow A$, onde essas ambigüidades de escopo não aparecem.

3 Lógicas Deônticas

As lógicas deônticas introduzem o símbolo primitivo O para 'é obrigatório que', a partir do qual os símbolos P para 'é permitido que' e F para 'é proibido que' são definidos: $PA = \sim O \sim A$ e $FA = O \sim A$. O análogo deôntico do axioma modal $(M): OA \rightarrow A$ é claramente não apropriado para a lógica deôntica. (Infelizmente, o que deve ser nem sempre é o caso). Entretanto, um sistema básico **D** de lógica deôntica pode ser interpretado adicionando o axioma mais fraco (D) a **K**.

$$(D) OA \rightarrow PA$$

O axioma (D) garante a consistência do sistema de obrigações insistindo que quando A é obrigatório, A é permissível. Um sistema que nos obriga a A , mas não nos permite fazê-lo, coloca-nos numa situação difícil. Apesar do fato de que alguns sustentarão que tais conflitos de obrigação são ao menos possíveis, a maioria dos lógicos deônticos aceita (D) .

$O(OA \rightarrow A)$ é outro axioma deôntico que parece desejável. Apesar de ser equivocado dizer que se A é obrigatório então A é o caso $(OA \rightarrow A)$, ainda assim, essa condicional *deveria* ser o caso. Alguns lógicos deônticos acreditam que D precisa ser suplementado com $O(OA \rightarrow A)$ também.

A controvérsia a respeito da iteração (repetição) de operadores surge novamente na lógica deôntica. Em algumas concepções de obrigação, OOA coincide com OA . 'deve ser o caso que deve ser o caso' é tratado como uma espécie de repetição desnecessária, os 'deve's' adicionais não adicionam nada de novo. Axiomas então são adicionados para garantir a equivalência entre OOA e OA . A política de iteração mais abrangente incorporada em **S5**

também pode ser adotada. Entretanto, há concepções de obrigação nas quais a distinção entre OA e OOA é preservada. A ideia é de que há diferença genuína entre as obrigações que nós *realmente* temos e as obrigações que nós *deveríamos* adotar. Então, por exemplo, 'deve ser o caso que deve ser o caso que A ' comanda a adoção de alguma obrigação que pode não estar de fato estabelecida, resultando no fato de que OOA pode ser verdadeiro mesmo quando OA é falso.

4 Lógicas Temporais

Na lógica temporal, há dois operadores básicos, G para o futuro, e H para o passado. G é lido 'será sempre o caso que' e o operador definido F (lido 'será o caso que') pode ser introduzido por $FA = \sim G\sim A$. De forma similar H é lido: 'sempre será o caso que' e P (para 'foi o caso que') é definido por $PA = \sim H\sim A$. Um sistema básico para lógica temporal chamado **Kt** resulta de adotar os princípios de **K** tanto para G quanto H , juntamente a dois axiomas para governar a interação entre os operadores do passado e do futuro:

Regras de "Necessitação":

Se A é um teorema então GA e $HÁ$ também o são.

Axiomas de Distribuição:

$$G(A \rightarrow B) \rightarrow (GA \rightarrow GB) \text{ e } H(A \rightarrow B) \rightarrow (HA \rightarrow HB)$$

Axiomas de Interação:

$$A \rightarrow GPA \text{ e } A \rightarrow HFA$$

Os axiomas de interação levantam questões a respeito das assimetrias entre o passado e o futuro. Uma intuição padrão é a de que o passado está fixo, enquanto o futuro ainda está em aberto. O primeiro axioma de interação ($A \rightarrow GPA$) está em conformidade com essa intuição em relatar que o que é o caso agora (A), irá em todos os tempos futuros, estar no passado (GPA). Entretanto, $A \rightarrow HFA$ pode parecer possuir caráter inaceitavelmente determinista, pois afirma, aparentemente, que o que é verdade agora (A) sempre foi de tal

forma que irá ocorrer no futuro (*HFA*). Entretanto, a semântica de mundos possíveis para a lógica temporal revela que essa preocupação resulta de uma simples confusão, e que os dois axiomas de interação são igualmente aceitáveis.

Note que o axioma característico da lógica modal, (*M*): $\Box A \rightarrow A$, não é aceitável nem para *H* nem para *G*, já que *A* não se segue de 'sempre foi o caso que *A*', nem de 'sempre será o caso que *A*'. Entretanto, é aceitável numa lógica temporal aparentada na qual *G* é lido como 'é e sempre será o caso que', e *H* é lido como 'é e sempre foi o caso que'.

Dependendo de quais suposições são feitas a respeito da estrutura do tempo, axiomas ulteriores precisam ser adicionados às lógicas temporais. Segue-se uma lista de axiomas comumente adotados nas lógicas temporais. Uma explicação de como eles dependem da estrutura do tempo será encontrada na seção Semântica de Mundos Possíveis.

$$GA \rightarrow GGA \text{ e } HA \rightarrow HHA$$

$$GGA \rightarrow GA \text{ e } HHA \rightarrow HA$$

$$GA \rightarrow HA \text{ e } HA \rightarrow PA$$

É interessante notar que certas combinações de operadores de passado e futuro podem ser utilizadas para expressar sentenças complexas na língua portuguesa. Por exemplo, *FPA* corresponde à sentença *A* no futuro do presente composto do indicativo, (como em 'daqui a 20 segundos a iluminação terá mudado'). Similarmente, *PPA* expressa o pretérito-mais-que-perfeito.

5 Lógicas Condicionais

O fundador da lógica modal, C. I. Lewis, definiu uma série de lógicas modais que não possuíam \Box como símbolo primitivo. Lewis estava ocupado em desenvolver uma lógica dos condicionais que estivesse livre dos chamados Paradoxos da Implicação Material, a saber os teoremas clássicos $A \rightarrow (\sim A \rightarrow B)$ e $B \rightarrow (A \rightarrow B)$. Ele introduziu o símbolo \supset para “implicação estrita” e desenvolveu lógicas onde nem $A \supset (\sim A \supset B)$ nem $B \supset (A \supset B)$ são demonstráveis. A prática moderna tem sido a de definir $A \supset B$ como $\Box(A \rightarrow B)$, e usar as lógicas modais governando \Box para obter resultados similares. Entretanto, a

demonstrabilidade de fórmulas tais como $(A \& \sim A) \rightarrow B$ em tais lógicas parece ir de encontro à preocupação com os paradoxos. Anderson e Belnap (1975) desenvolveram os sistemas **R** e **E** (para Lógica Relevante) que foram desenvolvidas para superar tais dificuldades. Esses sistemas requerem a revisão dos sistemas padrão de lógica proposicional. (Para uma discussão mais detalhada da Lógica de Relevância, veja o verbete Lógica Relevante).

David Lewis (1973) desenvolveu lógicas condicionais especiais para lidar com expressões contrafatuais, isso é, expressões da forma 'se A acontecesse, então B aconteceria'. (Kvart (1980) é outro ótimo recurso sobre o tópico). Lógicas contrafatuais diferem daquelas baseadas na implicação estrita porque as primeiras rejeitam enquanto as últimas aceitam a contraposição.

6 Semântica de Mundos Possíveis

O propósito da lógica é caracterizar a diferença entre argumentos válidos e argumentos inválidos. Um sistema lógico para uma linguagem é um conjunto de axiomas e regras designado para demonstrar *exatamente* os argumentos válidos enunciáveis na linguagem. Criar tal lógica pode ser uma tarefa difícil. O lógico precisa ter certeza de que o sistema é *correto*, isto é, que cada argumento demonstrado utilizando as regras e axiomas é de fato válido. Além disso, o sistema precisa ser *completo*, isto é, cada argumento válido deve possuir uma demonstração no sistema. Verificar a correção e completude de sistemas formais é uma preocupação central dos lógicos.

Tal verificação não pode ser realizada até que o conceito de validade seja definido rigorosamente. As semânticas formais para uma lógica fornecem uma definição de validade caracterizando o comportamento da verdade das sentenças do sistema. Na lógica proposicional, a validade pode ser definida utilizando tabelas de verdade. Um argumento válido é simplesmente um argumento em que cada linha da tabela de verdade que torna suas premissas verdadeiras também torna sua conclusão verdadeira. Entretanto, tabelas de verdade não podem ser utilizadas para fornecer uma explicação da validade nas lógicas modais porque não há tabelas de verdade para expressões como 'é necessário que', 'é obrigatório que', e similares. (O problema é que o valor de verdade de *A* não determina o valor de verdade de

$\Box A$. Por exemplo, quando A é 'Cães são cães', $\Box A$ é verdadeira, mas quando A é 'Cães são animais de estimação', $\Box A$ é falsa.) Não obstante, semânticas para lógicas modais podem ser definidas introduzindo mundos possíveis. Iremos ilustrar a semântica de mundos possíveis para uma lógica da necessidade contendo os símbolos \sim , \rightarrow e \Box . Em seguida iremos explicar como a mesma estratégia pode ser adaptada para outras lógicas da família modal.

Na lógica proposicional, uma valoração para sentenças atômicas (ou uma linha na tabela de verdade) atribui um valor de verdade (V ou F) para cada variável proposicional p . Daí os valores de verdade das sentenças complexas são calculados com tabelas de verdade. Nas semânticas modais, um conjunto M de mundos possíveis é introduzido. Uma valoração então dá um valor de verdade para cada variável proposicional *em cada mundo possível* de M . Isso quer dizer que o valor atribuído a p para o mundo m pode diferir do valor atribuído a p para outro mundo m' .

O valor de verdade da sentença atômica p no mundo m dado pela valoração v pode ser escrito como $v(p, m)$. Dada esta notação, os valores de verdade (V para verdadeiro, F para falso) de sentenças complexas da lógica modal para uma dada valoração v (e o membro m do conjunto de mundos M) pode ser definida pelas seguintes cláusulas de verdade. ('sse' abrevia 'se somente se'.)

$$(\sim) v(\sim A, m) = V \text{ sse } v(A, m) = F.$$

$$(\rightarrow) v(A \rightarrow B, m) = V \text{ sse } v(A, m) = F \text{ ou } v(B, m) = V.$$

$$(5) v(\Box A, m) = V \text{ sse para cada mundo } m' \text{ em } M, v(A, m') = V.$$

As cláusulas (\sim) e (\rightarrow) simplesmente descrevem o comportamento padrão da tabela de verdade para a negação e a implicação material, respectivamente. De acordo com (5), $\Box A$ é verdadeiro (em um mundo m) exatamente quando A é verdadeiro em *todos* os mundos possíveis. Dada a definição de \Diamond , (a saber, $\Diamond A = \sim \Box \sim A$) a condição de verdade (5) garante que $\Diamond A$ é verdadeiro no caso de A ser verdadeiro em *algum* mundo possível. Já que as cláusulas de verdade para \Box e \Diamond envolvem os quantificadores 'todo' e 'algum' (respectivamente), os paralelos entre o comportamento lógico entre \Box e $\forall x$, e entre \Diamond e $\exists x$ ressaltados na Seção 2, são esperados.

As cláusulas (\sim), (\rightarrow) e (5) nos permitem calcular o valor de verdade de qualquer sentença em qualquer mundo possível numa dada valoração. Uma definição de validade está agora muito próxima. Um argumento é *5-válido para um dado conjunto W* (de mundos possíveis) se e somente se cada valoração das sentenças atômicas que atribui V às premissas num mundo em M também atribuem V à conclusão no mesmo mundo. Um argumento é dito *5-válido* sse é válido para cada conjunto não-vazio M de mundos possíveis.

Mostrou-se que **S5** é correto e completo para *5-validade* (por isso nosso uso do símbolo '5'). Os argumentos 5-válidos são exatamente os argumentos demonstráveis em **S5**. Esse resultado sugere que **S5** é a forma correta de formular uma lógica da necessidade.

Entretanto, **S5** não é uma lógica razoável para todos os membros da família modal. Na lógica deôntica, na lógica temporal, e outras, o análogo da condição de verdade (5) claramente não é apropriado; além disso há até mesmo outras concepções de necessidade para as quais (5) também deveria ser rejeitada. O ponto é mais fácil de entender no caso da lógica temporal. Aqui, os membros de M são momentos do tempo, ou mundos “congelados”, por assim dizer, em um instante. Por questões de simplicidade vamos considerar uma lógica temporal *do futuro*, uma lógica onde $\Box A$ é lido '*será sempre o caso que*'. (Formulamos o sistema usando \Box ao invés do tradicional G para que as conexões com outras lógicas modais sejam mais facilmente apreciadas.) A cláusula correta para \Box deveria dizer que $\Box A$ é verdadeira no momento m sse A é verdadeiro em todos os momentos *no futuro de m* . Para restringir a atenção ao futuro, a relação R (precedência temporal) precisa ser introduzida. Assim, a cláusula correta pode ser formulada como se segue.

$$(K) v(\Box A, m) = V \text{ sse para cada } m', \text{ se } mRm', \text{ então } v(A, m') = V.$$

Isso diz que $\Box A$ é verdadeira em m somente no caso de A ser verdadeira em todos os momentos *depois de m* .

A validade para este tipo de lógica temporal pode agora ser definida. Um *enquadramento* $\langle M, R \rangle$ é um par consistindo de um conjunto não-vazio M (de mundos) e uma relação binária R em M . Um *modelo* $\langle E, v \rangle$ consiste num enquadramento E , e uma valoração v que atribui valores de verdade a cada sentença atômica em cada mundo de M . Dado um modelo, os valores de todas as sentenças complexas podem ser determinados utilizando (\sim), (\rightarrow) e (K). Um argumento é **K-válido** caso cada valoração das sentenças atômicas que atribui

V às premissas num mundo também atribui V à conclusão no mesmo mundo. Como o leitor pode ter adivinhado a partir do nosso uso de ‘**K**’, mostrou-se que a lógica modal mais simples **K** é tanto correta quanto completa para **K**-validade.

7 Axiomas Modais e Condições sobre Enquadramentos

Pode-se assumir nesta discussão que **K** é a lógica correta quando o \Box é lido como ‘será sempre o caso que’. Entretanto, há razões para pensar que **K** seja fraca demais. Uma característica lógica óbvia da relação R (pRecedência temporal) é a transitividade. Se wRv (w precede temporalmente v) e vRu (v precede temporalmente u), então se segue que wRu (w precede temporalmente u). Vamos definir então um novo tipo de validade que corresponde a essa condição sobre R . Seja um *4-modelo* qualquer modelo cujo enquadramento $\langle M, R \rangle$ é tal que R é uma relação transitiva em M . Então um argumento é 4-válido sse todo 4-modelo cuja valoração atribui V às premissas em um mundo também atribui V à conclusão no mesmo mundo. Usamos ‘4’ para descrever tal modelo transitivo porque a lógica que é adequada (tanto correta quanto completa) para 4-validade é **K4**, a lógica que resulta da adição do axioma (4): $\Box A \rightarrow \Box \Box A$ a **K**.

A transitividade não é a única propriedade que poderíamos querer exigir do enquadramento $\langle M, R \rangle$ se R é lido como ‘pRecedência temporal’ e M é um conjunto de momentos. Uma condição (que é apenas levemente controversa) é que não há um último momento no tempo, i.e. para cada mundo m há algum mundo n tal que mRn . Essa condição sobre os enquadramentos é chamada *serialidade*. A serialidade corresponde ao axioma (D): $\Box A \rightarrow \Diamond A$, da mesma maneira que a transitividade corresponde a (4). Um **D**-modelo é um **K**-modelo com um enquadramento serial. A partir do conceito de **D**-modelo a noção correspondente de **D**-validade pode ser definida, exatamente como fizemos no caso da 4-validade. Como você provavelmente adivinhou, o sistema que é adequado com relação à **D**-validade é **KD**, ou **K** adicionado de (D). Não apenas isso, mas o sistema **KD4** (isso é, **K** adicionado de (4) e (D)) é adequado a respeito da **D4**-validade, onde um **D4**-modelo é um modelo no qual $\langle M, R \rangle$ é serial e transitivo.

Outra propriedade que poderíamos desejar na relação de precedência temporal é a densidade, a condição que diz que entre dois momentos quaisquer, podemos sempre encontrar

outro. A densidade seria falsa se o tempo fosse atômico, isto é, se houvesse intervalos de tempo que não pudessem ser divididos em partes menores. A densidade corresponde ao axioma (C4): $\Box\Box A \rightarrow \Box A$, a recíproca de (4). Então, por exemplo, o sistema **KC4**, o qual corresponde a **K** adicionado de (C4), é adequado com relação aos modelos cujo enquadramento $\langle M, R \rangle$ é denso, e **KDC4**, adequado a respeito dos modelos cujos enquadramentos são seriais e densos, e assim por diante.

Cada um dos axiomas da lógica modal que discutimos anteriormente corresponde da mesma maneira a uma condição sobre os enquadramentos. O relacionamento entre condições sobre os enquadramentos e os axiomas correspondentes é um dos tópicos centrais no estudo das lógicas modais. Uma vez que a interpretação do operador intensional \Box foi decidida, as condições apropriadas sobre R podem ser determinadas para se ajustarem às correspondentes noções de validade. Por sua vez, isto nos permite selecionar o conjunto certo de axiomas para aquela lógica.

Por exemplo, considere uma lógica deôntica, na qual \Box é lido como ‘é obrigatório que’. Aqui a verdade de $\Box A$ não requer a verdade de A em *todos* os mundos possíveis, mas apenas num subconjunto desses mundos onde as pessoas fazem o que devem. Então também desejaremos introduzir a relação R para este tipo de lógica, e iremos utilizar a cláusula de verdade (K) para avaliar $\Box A$ em um mundo. Entretanto, nesse caso, R não é ‘precedência temporal’. Ao invés disso, mRm' vale caso o mundo m' seja uma variante moralmente aceitável de m , isto é, um mundo no qual nossas ações podem realizar o que é moralmente correto, ou certo, ou justo. Sob tal leitura, seria para ficar claro que os enquadramentos relevantes deveriam obedecer à serialidade, a condição que requer que cada mundo possível possua uma variante moralmente aceitável. A análise das propriedades desejadas para R torna claro que uma lógica deôntica básica pode ser formulada adicionando o axioma (D) a **K**.

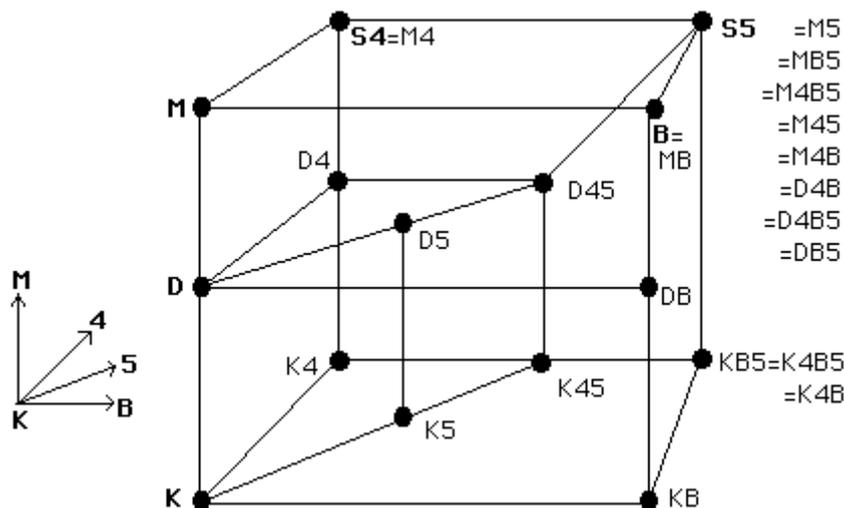
Mesmo na lógica modal, pode-se desejar restringir o leque de mundos possíveis que são relevantes para determinar se $\Box A$ é verdadeiro num dado mundo. Por exemplo, posso dizer que é necessário para mim pagar minhas contas, mesmo que eu saiba perfeitamente que há um mundo possível no qual eu falho em pagá-las. No discurso ordinário, a assertiva de que A é necessário não requer a verdade de A em *todos* os mundos possíveis, mas apenas numa subclasse de mundos que eu tenho em mente (por exemplo, mundos onde evito as sanções por falhar em pagar as contas). Para fornecer um tratamento genérico da necessidade, precisamos

dizer que $\Box A$ é verdadeira em m sse A é verdadeira em todos os mundos que *estão relacionados* a m da maneira adequada. Assim, para um operador \Box interpretado como necessidade, introduzimos a correspondente relação R no conjunto de mundos possíveis M , tradicionalmente chamada relação de acessibilidade. A relação de acessibilidade R vale entre os mundos m e m' sse m' é possível dados os fatos de m . Sob essa leitura de R , deve ficar claro que enquadramentos para lógicas modais deveriam ser reflexivos. Segue-se que as lógicas modais deveriam ser fundamentadas em **M**, o sistema que é resultado da adição de (M) a **K**. Dependendo de como a relação de acessibilidade é entendida, a simetria e a transitividade também podem ser desejáveis.

Uma lista de algumas das mais comumente discutidas condições sobre os enquadramentos e seus correspondentes axiomas, juntamente com um mapa mostrando a relação entre as várias lógicas modais pode ser encontrada na próxima seção.

8 Mapa das relações entre lógicas modais

O seguinte diagrama mostra o relacionamento entre as lógicas modais mais bem conhecida, a saber as lógicas que podem ser formadas adicionando alguma combinação dos axiomas (D) , (M) , (4) , (B) e (S) a **K**. Uma lista desses (e outros) axiomas, e das correspondentes condições sobre enquadramentos pode ser encontrada abaixo do diagrama.



Neste diagrama, os sistemas são dados pela lista de seus axiomas. Assim, por exemplo, **M4B** é o resultado de adicionar (*M*), (4) e (*B*) a **K**. Em negrito, indicamos os nomes tradicionais de alguns desses sistemas. Quando o sistema **S** aparece abaixo e/ou à esquerda de **S'** conectado por uma linha, então **S'** é uma extensão de **S**. Isso significa que cada argumento demonstrável em **S** é demonstrável em **S'**, mas **S** é mais fraco que **S'**, i.e. nem todos os argumentos demonstráveis em **S'** são demonstráveis em **S**.

A lista a seguir indica os axiomas, seus nomes, e as correspondentes condições sobre a relação de acessibilidade *R* para os axiomas até agora discutidos neste verbete da enciclopédia.

Nome	Axioma	Condição sobre os Enquadramentos	R é...
(<i>D</i>)	$\Box A \rightarrow \Diamond A$	$\exists u mRo$	Serial
(<i>M</i>)	$\Box A \rightarrow A$	mRm	Reflexiva
(4)	$\Box A \rightarrow \Box \Box A$	$(mRn \& nRo) \Rightarrow mRo$	Transitiva
(<i>B</i>)	$A \rightarrow \Box \Diamond A$	$mRn \Rightarrow nRm$	Simétrica
(5)	$\Diamond A \rightarrow \Box \Diamond A$	$(mRn \& mRo) \Rightarrow nRo$	Euclidiana
(<i>CD</i>)	$\Diamond A \rightarrow \Box A$	$(mRn \& mRo) \Rightarrow n=o$	Funcional
($\Box M$)	$\Box(\Box A \rightarrow A)$	$mRn \Rightarrow nRn$	Reflexiva Deslocada
(<i>C4</i>)	$\Box \Box A \rightarrow \Box A$	$mRn \Rightarrow \exists o(mRo \& oRn)$	Densa
(<i>C</i>)	$\Diamond \Box A \rightarrow \Box \Diamond A$	$mRn \& mRp \Rightarrow \exists o(nRo \& pRo)$	Convergente

Na lista de condições sobre os enquadramentos, as variáveis '*m*', '*n*', '*o*', '*p*' e o quantificador ' $\exists o$ ' têm como domínio *M*. '&' abrevia 'e' e ' \Rightarrow ' abrevia 'se...então'.

A noção de correspondência entre os axiomas e condições sobre enquadramentos que está em questão aqui foi explicada na seção anterior. Quando S é uma lista de axiomas e $F(S)$ é o conjunto correspondente de condições sobre enquadramentos, então S corresponde a $F(S)$ precisamente quando o sistema $K+S$ é adequado (correto e completo) para $F(S)$ -validade, isso é, um argumento é demonstrável em $K+S$ sse ele é $F(S)$ -válido. Várias noções mais fortes de correspondência entre axiomas e condições sobre enquadramentos emergiram da pesquisa sobre lógica modal.

9 O Axioma Geral

A correspondência entre axiomas e condições sobre enquadramentos pode parecer um mistério. Um belo resultado de Lemmon e Scott (1977) explica muito bem esse relacionamento. O teorema deles diz respeito aos axiomas que possuem a seguinte forma:

$$(G) \diamond^h \square^i A \rightarrow \square^j \diamond^k A$$

Utilizamos a notação ‘ \diamond^n ’ para representar uma cadeia de n losangos. Assim, por exemplo, ‘ \diamond^3 ’ abrevia uma cadeia de três losangos: ‘ $\diamond\diamond\diamond$ ’. De forma similar, ‘ \square^n ’ representa uma cadeia de n quadrados. Quando os valores de h, i, j e k são todos 1, temos o axioma (C):

$$(C) \diamond\square A \rightarrow \square\diamond A = \diamond^1\square^1 A \rightarrow \square^1\diamond^1 A$$

O axioma (B) resulta de atribuir 0 a h e a i , e atribuir 1 a j e a k :

$$(B) A \rightarrow \square\diamond A = \diamond^0\square^0 A \rightarrow \square^1\diamond^1 A$$

Para obter (4), podemos atribuir 0 a h e a k , atribuir 1 a i e 2 a j :

$$(4) \square A \rightarrow \square\square A = \diamond^0\square^1 A \rightarrow \square^2\diamond^0 A$$

Muitos (mas não todos) os axiomas da lógica modal podem ser obtidos dando os valores corretos aos parâmetros em (G).

Nossa próxima tarefa será dar as condições sobre enquadramentos que correspondem a (G) para uma dada seleção de valores para h, i, j e k . Para fazê-lo, vamos precisar de uma

definição. A composição de duas relações R e R' é uma relação $R \circ R'$ que é definida como se segue:

$$mR \circ R' v \text{ sse para algum } u, mRu \text{ e } uR'v.$$

Por exemplo, se R é a relação de ser um irmão, e R' é a relação de ser um pai ou mãe, então $R \circ R'$ é a relação de ser um tio ou tia, (pois m é o tio de v sse para alguma pessoa u , é o caso que m é o irmão de u e u é pai ou mãe de v). Uma relação pode ser composta consigo própria. Por exemplo, quando R é a relação de ser um pai ou mãe, então $R \circ R$ é a relação de ser um avô ou avó, e $R \circ R \circ R$ é a relação de ser um bisavô ou bisavó. Será útil escrever ' R^n ', para o resultado de compor R consigo própria n vezes. Então R^2 é $R \circ R$, e R^4 é $R \circ R \circ R \circ R$. Vamos considerar que R^1 é R e R^0 é a relação de identidade, isto é, $mR^0 v$ sse $m=v$.

Podemos agora enunciar o resultado de Scott-Lemmon. Ele nos diz que a condição sobre enquadramentos que corresponde exatamente a um axioma arbitrário da forma (G) é a seguinte.

$$(hijk\text{-Convergência}) mR^h v \ \& \ mR^i u \Rightarrow \exists x (vR^j x \ \& \ uR^k x)$$

É interessante notar como as condições familiares em R resultam de se atribuir valores para h, i, j e k de acordo com os valores no axioma correspondente. Por exemplo, considere (5). Nesse caso $i=0$, e $h=j=k=1$. Assim, a condição correspondente é

$$mRv \ \& \ mRu \Rightarrow \exists x (vR^0 x \ \& \ uRx).$$

Explicamos que R^0 é a relação de identidade. Assim, se $vR^0 x$ então $v=x$. Mas $\exists x (v=x \ \& \ uRx)$, é equivalente a uRv , e desta forma a condição Euclidiana é obtida:

$$(wRv \ \& \ wRu) \Rightarrow uRv$$

No caso do axioma (4), $h=0, i=1, j=2$ e $k=0$. Assim a correspondente condição sobre o enquadramento é

$$(m=v \ \& \ mR^2 u) \Rightarrow \exists x (vRx \ \& \ u=x).$$

Resolvendo as identidades obtemos:

$$vR^2 u \Rightarrow vRu$$

Pela definição de R^2 , vR^2u sse $\exists x (vRx \ \& \ xRu)$, então isso leva a:

$$\exists x(vRx \ \& \ xRu) \Rightarrow vRu,$$

o que pela lógica de predicados, é equivalente à transitividade.

$$vRx \ \& \ xRu \Rightarrow vRu.$$

O leitor pode achar um exercício apazível verificar como as condições correspondentes saem da *hijk*-Convergência quando os valores dos parâmetros *h*, *i*, *j* e *k* são dados por outros axiomas.

Os resultados de Scott-Lemmon fornecem um método rápido para estabelecer resultados sobre o relacionamento entre os axiomas e as correspondentes condições sobre os enquadramentos. Dado que eles mostraram a adequação de qualquer lógica que estende **K** pela adição de uma combinação de axiomas da forma (*G*) com relação aos modelos que satisfazem o conjunto correspondente de condições sobre os enquadramentos, eles forneceram demonstrações de adequação “no atacado” para a maioria dos sistemas na família modal. Sahlqvist (1975) descobriu generalizações importantes do resultado de Scott-Lemmon, cobrindo um escopo muito mais amplo de axiomas.

10 Lógicas da demonstrabilidade

A lógica modal foi útil para clarificar nosso entendimento de resultados centrais a respeito da demonstração nos fundamentos da matemática (Boolos, 1993). As lógicas da demonstrabilidade são sistemas onde as variáveis proposicionais *p*, *q*, *r*, etc. dizem respeito a fórmulas de algum sistema matemático, por exemplo, o sistema **PA** de Peano para a aritmética. (O sistema escolhido para a matemática pode variar, mas assuma que é **PA** para essa discussão). Gödel mostrou que a aritmética possui forte capacidade expressiva. Utilizando códigos numéricos para sentenças aritméticas, ele foi capaz de demonstrar uma correspondência entre sentenças da matemática e fatos sobre quais sentenças são e quais não são demonstráveis em **PA**. Por exemplo, ele demonstrou que há uma sentença *C* que é verdadeira exatamente quando nenhuma contradição é demonstrável em **PA** e há uma

sentença C (a famosa sentença de Gödel) que é verdadeira exatamente quando ela não é demonstrável em \mathbf{PA} .

Nas lógicas da demonstrabilidade, $\Box p$ é interpretada como uma fórmula (da aritmética) que expressa que o que p denota é demonstrável em \mathbf{PA} . Usando essa notação, as sentenças das lógicas da demonstrabilidade expressam fatos sobre demonstrabilidade. Suponha que $\Box\Box$ seja uma constante na lógica da demonstrabilidade denotando uma contradição. Então $\sim\Box\Box$ diz que \mathbf{PA} é consistente e $\Box A \rightarrow A$ diz que \mathbf{PA} é correto no sentido de que quando \mathbf{PA} demonstra A , A é de fato verdadeira. Por conseguinte, o quadrado pode ser iterado. Assim, por exemplo, $\Box\sim\Box\Box$ faz a dúvida a asserção de que \mathbf{PA} é capaz de demonstrar sua própria consistência, e $\sim\Box\Box \rightarrow \sim\Box\sim\Box\Box$ afirma (corretamente como Gödel provou) que se \mathbf{PA} é consistente então \mathbf{PA} é incapaz de demonstrar sua própria consistência.

Apesar das lógicas da demonstrabilidade formarem uma família de sistemas aparentados, o sistema \mathbf{GL} é de longe o mais conhecido. Ele resulta de adicionar o seguinte axioma a \mathbf{K} :

$$(GL) \Box(\Box A \rightarrow A) \rightarrow \Box A$$

O axioma (4): $\Box A \rightarrow \Box\Box A$ é demonstrável em \mathbf{GL} , portanto \mathbf{GL} é na verdade mais forte que $\mathbf{K4}$. Entretanto, axiomas como (M): $\Box A \rightarrow A$, e mesmo o mais fraco (D): $\Box A \rightarrow \Diamond A$ não estão disponíveis (nem são desejáveis) em \mathbf{GL} . Na lógica da demonstrabilidade, a demonstrabilidade não deve ser tratada como uma forma de necessidade. O motivo é que quando p é demonstrável em um sistema arbitrário para a matemática \mathbf{S} , não se segue que p seja verdadeiro, já que \mathbf{S} pode ser incorreto. Além disso, se p é demonstrável em \mathbf{S} ($\Box p$) não precisa nem mesmo seguir-se disso que $\sim p$ não possua uma demonstração ($\sim\Box\sim p = \Diamond p$). \mathbf{S} pode ser inconsistente e portanto demonstrar tanto p quanto $\sim p$.

O axioma (GL) captura o conteúdo do Teorema de Loeb, um resultado importante nos fundamentos da aritmética. $\Box A \rightarrow A$ diz que \mathbf{PA} é correto para A , isto é, que se A fosse demonstrada, A seria verdadeira. (Tal assertiva poderia não estar assegurada para um sistema \mathbf{S} arbitrariamente escolhido, já que A poderia ser demonstrável em \mathbf{S} e ser falsa.) (GL) afirma que se \mathbf{PA} consegue demonstrar a sentença que postula a correção para uma dada sentença A , então A já é demonstrável em \mathbf{PA} . O Teorema de Loeb denuncia um certo tipo de modéstia por parte de \mathbf{PA} (Boolos, 1993, p. 55). \mathbf{PA} nunca insiste (demonstra) que uma demonstração

de A implica na verdade de A , a não ser que **PA** já possua uma demonstração de A para dar suporte a tal afirmação.

Já foi demonstrado que **GL** é adequado para a demonstrabilidade no seguinte sentido. Digamos que uma sentença **GL** é *sempre demonstrável* exatamente quando a sentença da aritmética que ela denota é demonstrável, não importando quais valores são atribuídos às sentenças de **PA**. Então as sentenças demonstráveis de **GL** são exatamente as sentenças que são sempre demonstráveis. Esse resultado de adequação tem sido extremamente útil, já que questões gerais a respeito da demonstrabilidade em **PA** podem ser transformadas em questões mais fáceis sobre o que pode ser demonstrado em **GL**.

GL também pode ser caracterizada com uma semântica de mundos possíveis correta e completa. Uma condição correspondente sobre o enquadramento para caracterizar **GL**-validade é a de que o enquadramento seja transitivo, finito e irreflexivo.

11 Lógica Modal Avançada

As aplicações da lógica modal à matemática e à ciência da computação têm crescido em importância. A lógica da demonstrabilidade é apenas um exemplo dessa tendência. O termo “lógica modal avançada” se refere a uma tradição em pesquisa na lógica modal que é particularmente bem representada em departamentos de matemática e ciência da computação. Essa tradição tem se entrelaçado à história da lógica modal desde seu início (Goldblatt, 2006). A pesquisa sobre seus relacionamentos com topologia e álgebras representa alguns dos primeiros trabalhos técnicos em lógica modal. Entretanto, o termo ‘lógica modal avançada’ geralmente se refere a um segundo campo de trabalho desenvolvido desde a metade dos anos 70. Alguns exemplos dos muitos tópicos interessantes tratados neste campo incluem resultados sobre a decidibilidade (se é possível computar se uma fórmula de um dado sistema modal é um teorema) e complexidade (os custos em tempo e memória necessários para computar tais fatos sobre lógicas modais).

12 Bissimulação

A Bissimulação fornece um bom exemplo das relações frutíferas que foram desenvolvidas entre a lógica modal e a ciência da computação. Na ciência da computação, Sistemas de Transição Rotulados (STRs) são comumente utilizados para representar vias computacionais possíveis durante a execução de um programa. STRs são generalizações dos enquadramentos de Kripke, consistindo de um conjunto M de estados, e uma coleção de relações de i -acessibilidade R_i , uma para cada processo computacional i . Intuitivamente, $mRim'$ vale exatamente quando m' é um estado que resulta da aplicação do processo i ao estado m .

A linguagem da lógica polimodal ou dinâmica introduz uma coleção de operadores modais \Box_i , um para cada programa i (Harel, 1984). Nesse caso, $\Box_i A$ afirma que a sentença A vale em toda execução do processo i . Assim ideias como a corretude e a terminação bem-sucedida de programas podem ser expressadas nesta linguagem. Modelos para uma tal linguagem são como os modelos de Kripke, exceto que STRs são usados no lugar de enquadramentos. Uma *bissimulação* é uma relação de contraparte entre estados de dois desses modelos tal que exatamente as mesmas variáveis proposicionais são verdadeiras em estados contrapartes, e sempre que o mundo n é i -acessível a um dos estados contrapartes, então a outra contraparte possui a relação de i -acessibilidade a alguma contraparte de n . Em resumo, a estrutura de i -acessibilidade que alguém pode “ver” a partir de um dado estado imita o que se vê a partir de uma contraparte deste estado. A Bissimulação é uma noção mais fraca que o isomorfismo (uma relação de bissimulação não precisa ser bijetiva), mas é suficiente para garantir equivalência em termos de processamento.

Na década de 70, uma versão da bissimulação já havia sido desenvolvida por lógicos modais para ajudar a entender melhor o relacionamento entre axiomas de lógica modal e as condições correspondentes sobre os enquadramentos de Kripke. A semântica de Kripke fornece uma base para traduzir axiomas modais em sentenças de uma linguagem de segunda ordem na qual a quantificação é permitida sob letras predicativas P com um argumento. Substitua as metavaráveis A por sentenças abertas Px , traduza $\Box Px$ como $\Box y(Rxy \rightarrow Py)$, e feche as variáveis livres x e as letras predicativas unárias P com quantificadores universais. Por exemplo, a tradução na lógica de predicados do axioma esquemático $\Box A \rightarrow A$ é $\Box P \Box x [\Box y (Rxy \rightarrow Py) \rightarrow Px]$. Dada esta tradução, é possível instanciar a variável P por um predicado unário arbitrário, por exemplo o predicado Rx cuja extensão é o conjunto de todos os mundos

m tais que Rxm para um dado valor de x . Daí é possível obter $\Box x[\Box y(Rxy \rightarrow Rxy) \rightarrow Rxx]$, o que se reduz a $\Box xRxx$, já que $\Box y(Rxy \rightarrow Rxy)$ é uma tautologia. Isso ilumina a correspondência entre $\Box A \rightarrow A$ e a reflexividade dos enquadramentos ($\Box xRxx$). Resultados similares valem para muitos outros axiomas e condições de enquadramento. O “colapso” das condições axiomáticas de segunda-ordem em termos de condições de primeira-ordem sobre enquadramentos é muito útil para obter resultados de completude para lógicas modais. Por exemplo, essa é a ideia central por trás dos elegantes resultados de Sahlqvist (1975).

Mas quando é que a tradução em segunda ordem de um axioma se reduz desta maneira a uma condição de primeira ordem sobre R ? Na década de 70, van Benthem mostrou que isso acontece sse a validade da tradução em um modelo implica na sua validade em qualquer modelo bissimilar, onde dois modelos são bissimilares sse há uma bissimulação entre eles no caso especial no qual há uma única relação de acessibilidade. Esse resultado é generalizado facilmente para o caso polimodal (van Benthem 1996, p. 88). Isso sugere que a lógica polimodal está no nível exato de abstração para descrever, e raciocinar sobre, a computação e outros processos. (Afim de contas, o que realmente importa ali é a preservação dos valores de verdade das fórmulas em modelos, ao invés de detalhes mais sofisticados sobre a estrutura dos enquadramentos.) Além disso a tradução implícita dessas lógicas em termos de fragmentos bem compreendidos da lógica de predicados fornece uma riqueza de informações de interesse para a ciência da computação. Como resultado, uma frutífera área de pesquisa na ciência da computação tem sido desenvolvida tendo a bissimulação como sua ideia central (Ponse *et al.* 1995).

13 Quantificadores na Lógica Modal

Pareceria uma questão simples introduzir os quantificadores \Box (todo) e \Diamond (algum) na lógica modal. Poder-se-ia simplesmente adicionar as regras padrão (clássicas) para quantificadores aos princípios de qualquer lógica proposicional modal. Entretanto, a introdução de quantificadores em lógica modal envolve certas dificuldades. Algumas dessas são filosóficas. Por exemplo, Quine (1953) notoriamente sustentou que a quantificação em contextos modais é simplesmente incoerente, uma visão que originou uma gigantesca literatura. As reclamações de Quine já não carregam o peso que carregavam. (Veja Barcan

1990 para um bom resumo.) Não obstante, a visão de que há algo errado em “quantificar-se em” é ainda vastamente compartilhada. Um segundo tipo de complicação é técnico. Há uma grande variedade de escolhas que se pode fazer na semântica para lógicas modais quantificadas, e a demonstração de que um sistema de regras é apropriado para uma dada escolha pode ser difícil. Os trabalhos de Corsi (2002) e Garson (2005) avançam em direção a trazer unidade neste terreno, mas a situação permanece desafiadora.

Uma outra complicação é a de que alguns lógicos acreditam que a modalidade requer abandonar as regras de quantificação clássica em prol das regras mais fracas da lógica livre (Garson 2001). Os principais pontos de divergência no que diz respeito às regras de quantificação podem ser de fato ligados às decisões sobre como lidar com o domínio de quantificação. A alternativa mais simples, a abordagem do domínio fixo (às vezes chamada de possibilista), assume um domínio único de quantificação que contém todos os objetos possíveis. Por outro lado, a interpretação relativa ao mundo (ou atualista) assume que o domínio da quantificação muda de mundo para mundo, e contém apenas os objetos que são atuais em um dado mundo.

A abordagem do domínio fixo não requer grandes ajustes no maquinário clássico dos quantificadores. As lógicas modais que são adequadas para uma semântica de domínio fixo podem normalmente ser axiomatizadas adicionando princípios de uma lógica modal proposicional às regras clássicas de quantificação, juntamente com a Fórmula de Barcan (*BF*) (Barcan 1946). (Para uma explicação de algumas exceções interessantes ver Cresswell (1995)).

$$(BF) \Box x \Box A \rightarrow \Box \Box x A.$$

A interpretação do domínio fixo possui as vantagens da simplicidade e familiaridade, mas não fornece uma explicação direta da semântica de certas expressões quantificadas da linguagem natural. Nós não pensamos que ‘Existe um homem que assinou a Declaração de Independência dos EUA’ seja verdadeira, pelo menos não se lermos ‘existe’ no tempo presente. Não obstante, essa sentença foi verdadeira em 1777, o que mostra que o domínio para a expressão da linguagem natural ‘Existe um homem que’ muda para refletir quais homens existem em diferentes tempos. Um problema relacionado é o de que na interpretação do domínio fixo, a sentença $\Box y \Box \Box x(x=y)$ é válida. Entretanto, assumindo que $\Box x(x=y)$ seja

lido como: ‘y existe’, então $\Box y \Box x(x=y)$ diz que tudo existe necessariamente. No entanto, parece ser uma característica fundamental das ideias mais comuns sobre a modalidade que a existência de muitas coisas é contingente, e que diferentes objetos existem em diferentes mundos possíveis.

O defensor da interpretação do domínio fixo pode responder a essas objeções insistindo que em sua leitura dos quantificadores o domínio da quantificação contém *todos* os objetos possíveis, não apenas os objetos que existem num dado mundo. Assim o teorema $\Box y \Box x(x=y)$ faz a assertiva inócua de que todo objeto *possível* é necessariamente encontrado no domínio de todos os objetos possíveis. Além disso, as expressões quantificadas da linguagem natural cujo domínio é relativo ao mundo (ou momento) podem ser expressadas utilizando o quantificador de domínio fixo $\Box x$ e uma letra predicativa E com a leitura ‘atualmente existe’. Por exemplo, ao invés de traduzir ‘Existe algum Homem que Assinou a Declaração de Independência dos EUA’ por

$$\Box x(Hx \& Ax)$$

O defensor dos domínios fixos pode escrever:

$$\Box x(Ex \& Hx \& Ax),$$

assegurando assim que a tradução é falsa no tempo presente. Cresswell (1991) faz a interessante observação de que a quantificação relativa ao mundo possui poder expressivo limitado em relação à quantificação de domínio fixo. A quantificação relativa ao mundo pode ser definida com quantificadores de domínio fixo e E , mas não há maneira de expressar plenamente quantificadores de domínio fixo com quantificadores relativos ao mundo. Apesar disso argumentar a favor da abordagem clássica para a lógica modal quantificacional, a tática de tradução também aponta para algo como uma concessão em prol da lógica livre, pois os quantificadores relativos ao mundo assim definidos obedecem exatamente às regras da lógica livre.

Um problema com a estratégia de tradução utilizada por defensores da quantificação de domínio fixo é que interpretar a linguagem natural na lógica é algo menos direto, já que E precisa ser adicionado a todas as traduções de todas as sentenças cujas expressões quantificadas possuem domínios dependentes do contexto. Uma objeção mais séria à

quantificação de domínio fixo é que ela retira do quantificador um papel que Quine recomendou para ele, a saber, representar um comprometimento ontológico robusto. Nesta visão, o domínio de $\Box x$ precisa conter apenas entidades que são ontologicamente respeitáveis, e objetos possíveis são abstratos demais para assim merecer tal qualificação. Atualistas desta estirpe vão querer desenvolver a lógica de um quantificador $\Box x$ que reflete comprometimento com o que é atual em um dado mundo, ao invés do que é meramente possível.

Entretanto, os trabalhos recentes sobre o atualismo tendem a solapar tal objeção. Por exemplo, Linsky e Zalta (1994) sustentam que ao quantificador de domínio fixo pode ser dada uma interpretação que é perfeitamente aceitável aos atualistas. Atualistas que empregam semânticas de mundos possíveis rotineiramente quantificam sobre mundos possíveis em sua teoria semântica da linguagem. Pareceria então que mundos possíveis são atuais à luz do atualismo. Povoando o domínio com entidades abstratas não mais objetáveis que as aceitas pelos atualistas, Linsky e Zalta mostram que a Fórmula de Barcan e os princípios clássicos podem ser justificados. Note, entretanto, que atualistas podem responder que não estão comprometidos com a atualidade dos mundos possíveis dado que é entendido que os quantificadores utilizados em suas teorias da linguagem carecem de forte comprometimento ontológico. De qualquer forma, é uma questão aberta para os atualistas (e não atualistas também) investigar a lógica dos quantificadores com domínios mais robustos, por exemplo domínios excluindo mundos possíveis e outras entidades abstratas do gênero, e contendo apenas os particulares espaço-temporais encontrados num dado mundo. Para quantificadores desse tipo, domínios relativos ao mundo são apropriados.

Tais considerações motivam interesse nos sistemas que reconhecem a dependência do contexto na quantificação pela introdução de domínios relativos ao mundo. Aqui cada mundo possível possui seu próprio domínio de quantificação (o conjunto de objetos que atualmente existem no mundo), e os domínios variam de um mundo para o outro. Quando tal decisão é feita, surge uma dificuldade para a teoria clássica da quantificação. Note que a sentença $\Box x(x=t)$ é um teorema da lógica clássica, e portanto $\Box \Box x(x=t)$ é um teorema de **K** pela Regra da Necessitação. Suponha que o termo t denote Saul Kripke. Então o teorema anterior diz que é necessário que Saul Kripke exista, e portanto que ele está no domínio de cada mundo possível. Toda a motivação da abordagem relativa ao mundo era a de refletir a ideia de que objetos em um mundo podem não existir em outro mundo. Se as regras de quantificação

padrão são utilizadas, entretanto, cada termo t precisa se referir a algo que exista em todos os mundos possíveis. Isso parece incompatível com nossa prática ordinária de usar termos para nos referirmos a coisas que existem apenas contingentemente.

Uma resposta a essa dificuldade é simplesmente eliminar termos. Kripke (1963) deu um exemplo de um sistema que usa interpretação relativa ao mundo e preserva as regras clássicas. Entretanto, os custos são pesados. Primeiramente, a sua linguagem é artificialmente empobrecida, e em segundo lugar, as regras para a lógica modal proposicional precisam ser enfraquecidas.

Presumindo que desejamos uma linguagem que inclui termos, e que regras clássicas sejam adicionadas aos sistemas tradicionais da lógica modal proposicional, um novo problema surge. Em tal sistema, é possível provar (CBF), a recíproca da Fórmula de Barcan.

$$(CBF) \Box\Box xA \rightarrow \Box x\Box A$$

Esse fato possui sérias consequências para a semântica dos sistemas. Não é difícil mostrar que todo modelo relativo ao mundo de (CBF) precisa cumprir a condição (ND) (de ‘Domínios aninhados’).

(ND) Se mRn então o domínio de m é um subconjunto do domínio de n .

Entretanto (ND) entra em conflito com a ideia de introduzir domínios relativos ao mundo. A ideia afinal era de que a existência de objetos é contingente e portanto há mundos possíveis acessíveis onde algumas coisas no nosso mundo não existem.

Uma solução direta para esses problemas é abandonar as regras clássicas dos quantificadores e ao invés disso adotar regras para lógica livre (FL). As regras de FL coincidem com as regras clássicas, com exceção de que as inferências de $\Box xRx$ (tudo é atual) para Rp (Pégaso é atual) são bloqueadas. Isso é feito introduzindo o predicado ‘ E ’ (de ‘atualmente Existe’) e modificando a regra da instanciação universal. De $\Box xRx$ é permitido obter Rp apenas se também já se obteve Ep . Assumindo que o quantificador universal $\Box x$ é primitivo, e o quantificador existencial $\Box x$ é definido por $\Box xA =_{df} \sim \Box x\sim A$, então FL pode ser construída adicionando os dois princípios seguintes às regras da lógica proposicional

Generalização Universal. Se $B \rightarrow A(y)$ é um teorema, então $B \rightarrow \Box xA(x)$ também o é.

Instanciação Universal. $(\Box x A(x) \ \& \ En) \rightarrow A(n)$

(Aqui é assumido que $A(x)$ é uma fórmula bem formada arbitrária da lógica de predicados, e que $A(y)$ e $A(n)$ resultam de substituir cada ocorrência de x em $A(x)$ por y e n , de forma apropriada.) Note que o princípio da generalização universal é padrão, mas que o axioma da instanciação é restrito pela menção de En no antecedente. Em **FL**, as demonstrações de fórmulas como $\Box x \Box (x=t)$, $\Box y \Box \Box x (x=y)$, (CBF) , e (BF) , que parecem incompatíveis com a interpretação relativa ao mundo, são bloqueadas.

Uma objeção filosófica a **FL** é que E parece ser um predicado de existência, e muitos iriam sustentar que a existência não é uma propriedade legítima como ser verde ou pesar mais que dois quilos. Assim, os filósofos que rejeitam a ideia de que a existência é um predicado podem objetar a **FL**. Contudo, na maioria (mas não todas) as lógicas modais quantificadas que incluem identidade ($=$) essas preocupações podem ser contornadas definindo E como se segue.

$$Et \text{ =}_{df} \Box x (x=t)$$

A maneira mais geral de formular a lógica modal quantificada é criar **FS** adicionando as regras de **FL** a uma dada lógica modal proposicional **S**. Nas situações onde a quantificação clássica é desejada, pode-se simplesmente adicionar Et como um axioma a **FS**, de modo que os princípios clássicos tornam-se regras deriváveis. Os resultados de adequação para tais sistemas podem ser obtidos para a maioria das escolhas da lógica modal **S**, mas há exceções.

Uma complicação final na semântica para lógica modal quantificada merece ser mencionada. Ela surge quando expressões não-rígidas tais como ‘o inventor das bifocais’ são introduzidas na linguagem. Um termo é não-rígido quando se refere a diferentes objetos em diferentes mundos possíveis. O valor semântico de tal termo pode ser dado pelo que Carnap (1947) chamou de conceito individual, uma função que escolhe a denotação de um termo para cada mundo possível. Uma abordagem para lidar com termos não-rígidos é empregar a teoria de Russell das descrições definidas. Entretanto, em uma linguagem que trata expressões não-rígidas como termos genuínos, acontece que nem as regras de quantificação da lógica clássica, nem as da lógica livre são aceitáveis. (O problema não pode ser resolvido enfraquecendo a regra da substituição pela identidade.) Uma solução para esse problema é empregar um tratamento mais geral dos quantificadores, no qual o domínio da quantificação contenha

conceitos individuais ao invés de objetos. Essa interpretação mais geral fornece um ajuste melhor entre o tratamento dos termos e o tratamento dos quantificadores e resulta em sistemas que são adequados para regras clássicas ou regras de lógica livre (dependendo da escolha entre domínios fixos ou domínios relativos ao mundo).

Bibliografia

Uma excelente bibliografia de fontes históricas pode ser encontrada em Hughes e Cresswell (1968).

- Anderson, A. e Belnap, N., 1975, 1992, *Entailment: The Logic of Relevance and Necessity*, vol. 1 (1975), vol. 2 (1992), Princeton: Princeton University Press.
- Barcan (Marcus), R., 1947, "A Functional Calculus of First Order Based on Strict Implication," *Journal of Symbolic Logic*, 11: 1-16.
- Barcan (Marcus), R., 1967, "Essentialism in Modal Logic," *Noûs*, 1: 91-96.
- Barcan (Marcus), R., 1990, "A Backwards Look at Quine's Animadversions on Modalities," in R. Bartrett and R. Gibson (eds.), *Perspectives on Quine*, Cambridge: Blackwell.
- Bencivenga, E., 1986, "Free Logics," in D. Gabbay and F. Guentner (eds.), *Handbook of Philosophical Logic*, 3.6, Dordrecht: D. Reidel.
- Bentham, J. F. van, 1982, *The Logic of Time*, Dordrecht: D. Reidel.
- Bentham, J. F. van, 1983, *Modal Logic and Classical Logic*, Naples: Bibliopolis.
- Bentham, J. F. van, 1996, *Exploring Logical Dynamics*, Stanford: CSLI Publications.
- Blackburn, P., Rijke, M. de, and Venema, Y., 2001, *Modal Logic*, Cambridge: Cambridge University Press.
- Blackburn, P., Bentham, J. van, e Wolter, F., 2006, *Handbook of Modal Logic*, Amsterdam: North Holland.
- Bonevac, D., 1987, *Deduction*, Part II, Palo Alto: Mayfield Publishing Company.
- Boolos, G., 1993, *The Logic of Provability*, Cambridge: Cambridge University Press.
- Bull, R. e Segerberg, Krister, 1984, "Basic Modal Logic," in Gabbay, D., and Guentner, F. (eds.) *Handbook of Philosophical Logic*, 2.1, Dordrecht: D. Reidel.
- Carnap, R., 1947, *Meaning and Necessity*, Chicago: U. Chicago Press.
- Carnielli, W. e Pizzi, C., 2008, *Modalities and Multimodalities*, Heidelberg: Springer-Verlag.
- Chagrov, A. e Zakharyashev, M., 1997, *Modal Logic*, Oxford: Oxford University Press.
- Chellas, Brian, 1980, *Modal Logic: An Introduction*, Cambridge: Cambridge University Press.

- Cresswell, M. J., 2001, "Modal Logic", in L. Goble (ed.), *The Blackwell Guide to Philosophical Logic*, Oxford: Blackwell, 136-158.
- Cresswell, M. J., 1995, "Incompleteness and the Barcan formula", *Journal of Philosophical Logic*, 24: 379-403.
- Cresswell, M. J., 1991, "In Defence of the Barcan Formula," *Logique et Analyse*, 135-136: 271-282.
- Corsi, G, 2002, "A Unified Completeness Theorem for Quantified Modal Logics," *Journal of Symbolic Logic*, 67: 1483-1510.
- Fitting, M. e Mendelsohn, R., 1998, *First Order Modal Logic*, Dordrecht: Kluwer.
- Gabbay, D., 1976, *Investigations in Modal and Tense Logics*, Dordrecht: D. Reidel.
- Gabbay, D., 1994, *Temporal Logic: Mathematical Foundations and Computational Aspects*, New York: Oxford University Press.
- Garson, J., 2001, "Quantification in Modal Logic," in Gabbay, D., and Guentner, F. (eds.) *Handbook of Philosophical Logic*, second edition, volume 3, Dordrecht: D. Reidel, 267-323.
- Garson, J., 2005, "Unifying Quantified Modal Logic," *Journal of Philosophical Logic*, 34: 621-649.
- Garson, J., 2006, *Modal Logic for Philosophers*, Cambridge: Cambridge University Press.
- Goldblatt, R., 1993, *Mathematics of Modality*, CSLI Lecture Notes #43, Chicago: University of Chicago Press.
- Goldblatt, R., 2006, "Mathematical Modal Logic: a View of its Evolution," in D. Gabbay and J. Woods (eds.), *Handbook of the History of Logic*, vol. 6, Amsterdam: Elsevier.
- Harel, D., 1984, "Dynamic Logic," in D. Gabbay and F. Guentner(eds.), *Handbook of Philosophical Logic*, volume 2, Dordrecht: D. Reidel, 497-604.
- Hintikka, J., 1962, *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Ithaca, NY: Cornell University Press.
- Hilpinen, R., 1971, *Deontic Logic: Introductory and Systematic Readings*, Dordrecht: D. Reidel.
- Hughes, G. e Cresswell, M., 1968, *An Introduction to Modal Logic*, London: Methuen.
- Hughes, G. e Cresswell, M., 1984, *A Companion to Modal Logic*, London: Methuen.
- Hughes, G. e Cresswell, M., 1996, *A New Introduction to Modal Logic*, London: Routledge.
- Kripke, Saul, 1963, "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica*, 16: 83-94.
- Konyndik, K., 1986, *Introductory Modal Logic*, Notre Dame: University of Notre Dame Press.
- Kvart, I., 1986, *A Theory of Counterfactuals*, Indianapolis: Hackett Publishing Company.

- Lemmon, E. e Scott, D., 1977, *An Introduction to Modal Logic*, Oxford: Blackwell.
- Lewis, C.I. e Langford, C. H., 1959 (1932), *Symbolic Logic*, New York: Dover Publications.
- Lewis, D., 1973, *Counterfactuals*, Cambridge, MA: Harvard University Press.
- Linsky, B. e Zalta, E., 1994, "In Defense of the Simplest Quantified Modal Logic," *Philosophical Perspectives*, (Logic and Language), 8: 431-458.
- Mints, G. 1992, *A Short Introduction to Modal Logic*, Chicago: University of Chicago Press.
- Ponse, A., de Rijke, M., e Venema, Y., 1995, *Modal Logic and Process Algebra, A Bisimulation Perspective*, Stanford: CSLI Publications.
- Popkorn, S., 1995, *First Steps in Modal Logic*, Cambridge: Cambridge University Press.
- Prior, A. N., 1957, *Time and Modality*, Oxford: Clarendon Press.
- Prior, A. N., 1967, *Past, Present and Future*, Oxford: Clarendon Press.
- Quine, W. V. O., 1953, "Reference and Modality", in *From a Logical Point of View*, Cambridge, MA: Harvard University Press. 139-159.
- Rescher, N, e Urquhart, A., 1971, *Temporal Logic*, New York: Springer Verlag.
- Sahlqvist, H., 1975, "Completeness and Correspondence in First and Second Order Semantics for Modal Logic," in S. Kanger (ed.), *Proceedings of the Third Scandinavian Logic Symposium*, Amsterdam: North Holland. 110-143.
- Zeman, J., 1973, *Modal Logic, The Lewis-Modal Systems*, Oxford: Oxford University Press.

METODOLOGIA: OS ELEMENTOS DA FILOSOFIA DA CIÊNCIA*

David Papineau

*Tradução de Luiz Helvécio Marques Segundo¹***INTRODUÇÃO**

O conteúdo da *metodologia* é melhor definido em oposição ao da *lógica*. A lógica é o estudo do raciocínio dedutivamente válido: num argumento dedutivamente válido as premissas fornecem razões conclusivas para a conclusão; é completamente impossível que as premissas sejam verdadeiras e a conclusão falsa. No entanto, grande parte do raciocínio que efetivamente usamos não corresponde a esse ideal. Tanto na vida comum quanto na ciência, os argumentos que usamos não fornecem razões conclusivas para as suas conclusões. Eles podem, em algum sentido, nos fornecer boas razões para acreditar em suas conclusões, mas não nos obriga da mesma maneira absoluta que os argumentos dedutivos.

A seguinte discussão acerca da metodologia tratará desse tipo de raciocínio não-dedutivo e das várias questões filosóficas que surgem na tentativa de entendê-lo. Haverá cinco seções: (1) A Indução e seus Problemas; (2) Leis da Natureza; (3) Realismo, Instrumentalismo e Subdeterminação; (4) Confirmação e Probabilidade; (5) Explicação.

1. A INDUÇÃO E SEUS PROBLEMAS**O problema da indução**

Em geral, “indução” se refere a qualquer forma de inferência na qual passamos de um conjunto finito de resultados observacionais ou experimentais a uma conclusão sobre como as coisas geralmente se comportam. Há várias formas de inferência indutiva, mas nos concentraremos nas *induções enumerativas* simples, que partem da premissa de

* Tradução de “Methodology: The elements of the philosophy of science” de David Papineau. In: *Philosophy 1: A Guide Through the Subject*, org. A. C. Grayling; Oxford University Press, 1995.

¹ Revisor: Rodrigo Reis Lastra Cid.

que um fenômeno até agora tem se seguido de outro, e concluem que esses fenômenos sempre ocorrerão simultaneamente. Por exemplo, você poderia notar que em todas as vezes em que viste o céu vermelho ao entardecer o tempo foi bom no dia seguinte, e concluir com base nisso que o céu vermelho ao entardecer é sempre seguido por bom tempo. Ou você poderia notar que todas as amostras de sódio aquecidas num bico de Bunsen produziram um brilho alaranjado, e concluir com base nisso que em geral *todo* sódio aquecido produz um brilho alaranjado. Esquemáticamente, a premissa de uma indução enumerativa é “n As observados foram Bs”, e a conclusão é “Todos os As são Bs”.

Note que essas inferências indutivas começam com premissas particulares sobre um número finito de observações do passado, mas terminam com uma conclusão geral sobre como a natureza *sempre* se comportará. Essa é a fonte do notório *problema da indução*. Pois não é claro como qualquer quantidade finita de informação sobre o que aconteceu no passado possa garantir que um padrão natural continuará por todo o tempo.

Afinal, o que exclui a possibilidade do curso da natureza poder mudar e os padrões que observamos até agora se tornarem um guia fraco para o futuro? Ainda que, até agora, todos os finais de tarde com o céu avermelhado tenham sido seguidos por bom tempo, quem disse que não começarão a ser seguidos por chuva no próximo século? Ainda que todo o sódio aquecido até agora tenha brilho alaranjado, quem disse que não começará a produzir uma chama azulada em algum tempo futuro?

Nesse aspecto a indução contrasta com a dedução. Nas inferências dedutivas as premissas realmente garantem a conclusão. Por exemplo, se você sabe que “Ou esta substância é sódio ou é potássio”, e depois descobre que “Não é sódio”, pode concluir com certeza que “É potássio”. A verdade das premissas não deixa espaço para a conclusão ser algo além de verdadeira. Mas isso não acontece numa inferência indutiva. Se você disse que “Cada um dos As observados até agora foi B”, isso não garante que “Todos os As, incluindo os futuros, são Bs”. É perfeitamente possível que a primeira afirmação seja verdadeira e a última falsa.

Ilustrei o problema da indução no que diz respeito às induções enumerativas. Há outra forma de indução além da indução enumerativa, como veremos mais adiante. Mas

o problema da indução é bastante geral. Pois aquilo que as diferentes formas de indução têm em comum é que nos levam de informação sobre uma quantidade finita de instâncias a uma conclusão mais geral sobre uma classe mais ampla de casos. Uma vez que nada na lógica parece garantir que a classe mais ampla apresentará o mesmo comportamento que as instâncias finitas, qualquer inferência desse tipo é por essa razão igualmente problemática.

O problema da indução ameaça tanto o conhecimento cotidiano quanto o científico. Grande parte do conhecimento cotidiano em que confiamos consiste em princípios gerais como “*Sempre* que você se corta, sangra”, ou “*Sempre* que os freios são puxados, o carro pára”. Similarmente, todas as descobertas científicas que merecem esse nome estão na forma de princípios gerais: a lei de Galileu da queda livre diz que “Todos os corpos caem com aceleração constante”; a lei de Newton da gravitação diz que “Todos os corpos se atraem na proporção de suas massas e na proporção inversa do quadrado da distância entre eles”; a lei de Avogrado diz que “Todos os gases com a mesma temperatura e pressão contêm o mesmo número de moléculas por unidade de volume”; e assim por diante. O problema da indução põe em xeque a autoridade dessas afirmações gerais. Pois se o nosso indício é simplesmente o de que essas generalizações têm funcionado até agora, então como podemos estar certos de que elas não serão infirmadas por ocorrências futuras?

Respostas iniciais ao problema

Um princípio de indução

Uma possível resposta ao problema da indução seria apelar para algum “princípio de indução” que diz que, para uma quantidade N ,

(P) Para quaisquer α e β , sempre que N α s observados são β s, então todos os α s são β s.

Se tal princípio estivesse disponível, então poderíamos adicioná-lo à premissa original de qualquer indução enumerativa – a saber, que N (ou mais) A s observados são B s – de modo a concluir dedutivamente que “Todos os A s são B s”. Pois uma vez que

adicionarmos (P) como uma premissa, então já não há mais espaço para que as premissas da indução sejam verdadeiras e a conclusão falsa.

No entanto, ainda que deixemos de lado a questão do quão grande N precisa ser para tornar (P) plausível, há uma dificuldade óbvia sobre o estatuto do princípio proposto. (P) claramente não é uma afirmação analítica cuja verdade é garantida por seu significado: você poderia entender todos os seus termos mesmo sem acreditar nela. Portanto, ela tem de ser uma afirmação sintética e que precisa de apoio de índices empíricos. Mas uma vez que (P) é uma generalização, esse apoio teria de ser algum tipo de argumento indutivo, tomando como premissas algum corpo finito de instâncias nas quais as inferências funcionaram no passado, e procurando passar a (P) como conclusão. No presente contexto do argumento, portanto, isso seria pressupor aquilo que está em questão, que é defender argumentos indutivos contra o desafio levantado pelo problema da indução.

Argumentos indutivos a favor da indução

Suponha que abandonemos qualquer princípio de indução e por isso aceitamos que não podemos tornar dedutivos argumentos indutivos. Não podemos, contudo, argumentar simplesmente que os argumentos indutivos são aceitáveis porque *funcionam*? Ainda que as premissas não *garantam* logicamente as conclusões, elas normalmente não se mostram verdadeiras? Afinal, a nossa experiência não tem nos mostrado que padrões como céu avermelhado/bom tempo ou sódio aquecido/chama alaranjada continuam vigorando no futuro assim como se apresentaram no passado?

Mas essa sugestão incorre na mesma dificuldade que a anterior. Estamos argumentando que as induções são geralmente aceitáveis porque a nossa experiência tem mostrado que elas funcionaram até agora. Mas isso é em si um argumento indutivo. Afinal, ainda que os padrões observados até agora tenham vigorado, o que garante que continuarão a vigorar? Como Bertrand Russell disse uma vez, não adianta observar que os futuros do passado estiveram de acordo com os passados do passado; o que queremos saber é se os futuros do futuro estarão de acordo com os passados do futuro. Dado que estamos tentando defender a indução de objeções, um argumento indutivo a favor da indução mais uma vez pressupõe o que está em causa.

Introduzindo a probabilidade

Outra resposta possível ao problema da indução é considerar que as inferências indutivas geram apenas conclusões *prováveis*, ao invés de conclusões certas. Ainda que os indícios do passado não nos assegurem os padrões futuros, não poderiam pelo menos sustentar as conclusões sobre os padrões *prováveis*?

Mais adiante veremos que a idéia de probabilidade é de fato importante para o nosso entendimento dos argumentos indutivos. Mas não é difícil mostrar que por si própria ela não é suficiente para resolver o problema da indução.

De fato, com veremos mais adiante, há na verdade duas noções de probabilidade. Grosso modo, precisamos distinguir a probabilidade no sentido de *grau de crença* racional da probabilidade no sentido de *tendência objetiva*. Quando dizemos que a probabilidade de nevar hoje é de 50 por cento, poderíamos querer dizer uma de duas coisas. Primeiro, poderíamos estar expressando um grau de crença: dizendo que temos uma expectativa igual a de que nevará ou não nevará hoje. Alternativamente, poderíamos estar fazendo uma afirmação sobre uma tendência: dizendo que, em geral neva em 50 por cento dos dias como os de hoje. Mais adiante veremos em maior detalhe essas interpretações “subjetiva” e “objetiva” da probabilidade. Quero aqui apenas mostrar que nenhuma delas nos ajuda no problema da indução.

Suponha primeiro que a conclusão de uma inferência indutiva é um enunciado de probabilidade *objetiva*, estabelecendo, digamos, que em 90 por cento dos casos em que os As se mostraram Bs (por exemplo, que em 90 por cento dos dias seguidos de entardeceres com o céu avermelhado há bom tempo). Os indícios para essa afirmação serão, contudo, um corpo finito de observações, a saber, que em nossa experiência *até agora* mais ou menos 90 por cento de As foram BS. Portanto, o problema da indução está ainda conosco, pois ainda precisamos explicar como um corpo finito de indícios pode estabelecer uma conclusão geral. Pois, note que a conclusão probabilística é ainda uma afirmação alegando não apenas que 90 por cento de As foram Bs no passado, mas também que isso continuará no futuro. Ainda que o padrão no qual estejamos agora interessados seja probabilístico, mesmo sem exceção, ainda enfrentamos a mesma

dificuldade em explicar como os padrões do passado podem nos dizer algo sobre os do futuro.

Alternativamente, poderíamos tomar a conclusão de uma inferência indutiva como sendo um enunciado de *probabilidade* subjetiva, declarando que “Deveríamos atribuir um alto grau de crença à proposição de todos os As são Bs”. (Note que poderíamos também ter um enunciado de probabilidade subjetiva sobre uma proposição de probabilidade objetiva: por exemplo, “Deveríamos atribuir um alto grau à proposição de que 90 por cento de As são Bs”). A dificuldade, mais uma vez, é que nossos indícios para tal conclusão sobre a probabilidade subjetiva é simplesmente que As foram observados junto de Bs *até agora*. A conclusão, contudo, diz que deveríamos ter uma expectativa alta de que os As estarão junto dos Bs no futuro assim como no passado. Portanto, ainda enfrentamos o problema de explicar como fatos sobre o passado podem nos dizer o que pensar sobre o futuro.

A alternativa de Popper à indução

Deve-se a Karl Popper uma linha de resposta bastante diferente ao problema da indução. Popper olha para a prática da ciência a fim de nos mostrar como lidar com o problema. Na perspectiva de Popper, a ciência, em primeiro lugar, não repousa na indução. Ele nega que os cientistas comecem com observações e depois inferem uma teoria geral. Ao invés, eles primeiro avançam uma teoria como uma conjectura inicialmente não corroborada, e então comparam suas previsões com as observações para verem se resistem ao teste. Se tais testes se provarem negativos, então a teoria é experimentalmente falseada e os cientistas procurarão alguma alternativa nova. Se, por outro lado, os testes se enquadrarem na teoria, então os cientistas continuarão a sustentando – não como uma verdade admitidamente demonstrada, mas como uma conjectura não derrubada.

Se olharmos para a ciência desse modo, argumenta Popper, então veremos que ela não precisa da indução. De acordo como Popper, as inferências que importam para a ciência são as *refutações*, que toma como premissa uma previsão malograda e conclui que a teoria por trás dessa previsão é falsa. Essas inferências não são indutivas, mas antes dedutivas. Vemos que algum A é um não-B, e concluímos que não é o caso que

todos os As sejam Bs. Não há espaço aqui para que as premissas sejam verdadeiras e a conclusão falsa. Se descobrirmos que alguma amostra de sódio não fica alaranjada quando aquecida, então sabemos com certeza que não é o caso que todo sódio aquecido fica alaranjado. O ponto aqui é que é muito mais fácil refutar teorias do que prová-las. Um único exemplo contrário constituirá uma prova conclusiva.

De acordo com Popper, a ciência é, portanto, uma seqüência de conjecturas. As teorias científicas são desenvolvidas como hipóteses e são substituídas por novas hipóteses quando são falseadas. No entanto, essa perspectiva da ciência levanta uma pergunta óbvia: se as teorias científicas são sempre conjecturais, então o que torna a ciência melhor do que a astrologia, ou o espiritismo, ou qualquer outra forma de superstição injustificada? Um não-poperiano responderia essa pergunta dizendo que a verdadeira ciência prova suas alegações baseando-se em indícios observacionais, ao passo que a superstição não é nada além de mera conjectura. Mas na abordagem de Popper, mesmo as teorias científicas são mera conjectura – pois não podem ser provadas pelas observações, sendo apenas conjecturas não derrubadas.

Popper chama isso de problema da demarcação – qual a diferença entre ciência e outras formas de crença? A sua resposta é que a ciência, ao contrário da superstição, é pelo menos *falseável*, ainda que não seja demonstrável. As teorias científicas são concebidas em termos precisos, e por isso produzem previsões definidas. Por exemplo, as leis de Newton nos dizem exatamente onde certos planetas aparecerão em certos momentos. E isso significa que se tais previsões falharem, podemos ter certeza de que a teoria por trás delas é falsa. Em contraste, sistemas de crenças como a astrologia são irremediavelmente vagos, de modo que evitam sempre que se mostre que estão definitivamente errados. A astrologia pode prever que um escorpiano prosperará em suas relações pessoais nas quintas-feiras, mas ao se deparar com um escorpiano cuja esposa foi embora na quinta-feira, os defensores da astrologia provavelmente responderão que no fim das contas o fim do casamento foi o melhor. Por causa disso, nada forçará os astrólogos a admitir que a teoria deles esteja errada. A teoria é expressa em termos tão imprecisos que nenhuma observação efetiva poderão falsificá-la.

O próprio Popper usa esse critério de falseabilidade para distinguir a ciência genuína, não apenas dos sistemas tradicionais de crença como a astrologia e o espiritismo, mas também do marxismo, da psicanálise, e de várias outras disciplinas contemporâneas que ele designa como pseudociências. De acordo com Popper, as afirmações centrais dessas teorias são tão não-falseáveis quanto as da astrologia. Os marxistas prevêem que as revoluções do proletariado serão bem sucedidas quando os regimes capitalistas estiverem suficientemente enfraquecidos por suas contradições internas. Mas ao se deparar com revoluções proletárias mal sucedidas, eles simplesmente respondem que as contradições nesses regimes capitalistas particulares ainda não tinham o enfraquecido suficientemente. Similarmente, as teorias psicanalíticas alegarão que todas as neuroses dos adultos se devem a traumas de infância, mas quando se deparam com adultos perturbados com infâncias aparentemente não perturbadas, dirão que aqueles adultos sofreram traumas psicológicos privados quando jovens. Os cientistas genuínos dirão de antemão quais descobertas observacionais os fariam mudar de idéia, e abandonariam suas teorias se essas descobertas fossem feitas. Mas os teóricos marxistas e psicanalistas concebem suas teorias de tal modo que, argumenta Popper, nenhuma observação os farão ajustar as suas ideias.

As falhas do falseacionismo

À primeira vista, Popper parece oferecer uma maneira atrativa de lidar com o problema da indução. Não há, contudo, razão para duvidar que ele de fato ofereceu uma solução.

A objeção central à sua abordagem é que ela apenas dá conta do conhecimento científico *negativo*, oposto ao conhecimento *positivo*. Popper chama a atenção que um único contra-exemplo pode nos mostrar que uma teoria científica está errada. Mas ele não diz qualquer coisa sobre o que pode nos mostrar que uma teoria científica está correta. Contudo, é o conhecimento positivo desse último tipo que supostamente se segue das inferências indutivas. Além do mais, é esse tipo de conhecimento positivo que torna a indução tão importante. Podemos curar doenças e mandar pessoas à lua porque sabemos que certas causas sempre têm certos resultados, não porque sabemos que *não* os têm. Se Popper não pode explicar como às vezes sabemos que “Todos os As são Bs”, ao

invés de apenas “É falso que todos os As são Bs”, então ele certamente falhou em lidar de maneira adequada como o problema da indução.

A resposta comum de Popper a essa objeção é que ele está preocupado como a lógica da pesquisa científica pura, e não com questões práticas sobre aplicações tecnológicas. A pesquisa científica requer apenas que formulemos conjecturas falseáveis e as rejeitemos caso descubramos contraexemplos. A questão adicional de saber se deveríamos acreditar naquelas conjecturas e confiar em suas previsões quando, digamos, prescrevemos alguma droga ou construímos um dique, Popper considera como uma questão essencialmente prática, e como tal não faz parte da análise da prática científica racional.

Mas isso não será assim caso Popper esteja supostamente oferecendo uma solução ao problema da indução. O problema da indução é essencialmente o problema de como podemos basear juízos sobre o futuro a partir de indícios sobre o passado. Ao insistir que as teorias científicas são apenas conjecturas, e que, portanto, não temos base racional pra acreditar em suas previsões, Popper está simplesmente negando que possamos fazer juízos racionais sobre o futuro.

Considere as duas previsões:

Quando eu pular dessa janela do décimo andar, vou dolorosamente me espatifar no chão.

Quando eu pular da janela, flutuarei como uma pena até aterrissar suavemente.

Intuitivamente é mais racional acreditar em (A), que presume que o futuro será como o passado, do que (B), que não o faz. Mas Popper, uma vez que rejeita a indução, está comprometido com a idéia de que os indícios do passado não tornam quaisquer crenças sobre o futuro mais racionais do que quaisquer outras e, portanto, com a idéia de que acreditar em (B) não é menos racional que acreditar em (A).

Algo vai mal. Acreditar em (A) é, *com certeza*, mais racional do que acreditar em (B). Ao dizer isso, não quero negar que haja um *problema* da indução. Na verdade, é precisamente *porque* acreditar em (A) é mais racional do que acreditar em (B) que a

indução é problemática. Todo mundo, exceto Popper, pode ver que acreditar em (A) é mais racional que acreditar em (B). O problema é então explicar, frente ao fato de a indução não ser logicamente garantida, *por que* acreditar em (A) é mais racional do que acreditar em (B). Portanto, a negação de Popper da superioridade racional de (A) sobre (B) não é tanto uma solução ao problema da indução, mas simplesmente a recusa de reconhecer o problema em primeiro lugar. Como um crítico de um dos livros de Popper uma vez disse, a atitude de Popper com a indução é como a de alguém que se põe na linha de partida de uma corrida e grita, “Venci, venci”.

Ainda que falhe em lidar com a indução, dever-se-ia reconhecer que a filosofia da ciência de Popper tem alguma força enquanto descrição da pesquisa científica pura. Pois certamente é verdade que muitas teorias científicas nascem como conjecturas, exatamente como Popper descreve. Por exemplo, quando a teoria de Einstein da relatividade geral foi pela primeira vez proposta, poucos cientistas acreditavam efetivamente nela. Ao invés disso, consideravam-na como uma hipótese interessante e estavam curiosos para ver se era verdadeira. Nesse estágio inicial da vida de uma teoria, as recomendações de Popper fazem um sentido eminente. Obviamente que se você quiser ver se uma teoria é verdadeira, o passo seguinte é submetê-la ao teste observacional. E para esse propósito é importante que a teoria seja concebida em termos bastante precisos a fim de que os cientistas testem aquilo que ela implica sobre o mundo observável – isto é, em termos bastante precisos para ser falseável. E naturalmente, se a nova teoria for falseada, então os cientistas a rejeitarão e procurarão alguma alternativa, ao passo que se suas previsões forem corroboradas, então os cientistas continuarão a investigá-la.

Onde a filosofia da ciência de Popper corre mal, no entanto, é ao sustentar que as teorias científicas nunca progridem além do nível da conjectura. Como já consenti, as teorias são geralmente meras conjecturas quando são pela primeira vez desenvolvidas, e podem permanecer como conjecturas até que os primeiros indícios apareçam. Mas em muitos casos o acúmulo de indícios em favor de uma teoria as promoverá do estatuto de conjectura ao de verdade estabelecida. A teoria geral da relatividade nasceu como uma conjectura, e ainda assim muitos cientistas a consideraram hipotética, mesmo após as famosas observações iniciais de Sir Arthur Eddington em 1919 da luz aparentemente se

curvando próximo ao sol. Mas nesse meio tempo esse indício inicial foi suplementado com indícios na forma de desvios gravitacionais para o vermelho, dilatação do tempo, e buracos negros; seria considerado um cientista excêntrico aquele que hoje considerasse a teoria geral da relatividade como menos do que firmemente estabelecida.

Esse exemplo pode ser multiplicado. A teoria heliocêntrica do sistema solar, a teoria da evolução por seleção natural, e a teoria da deriva continental nasceram como conjecturas intrigantes e com poucos indícios a favor delas sobre suas rivais. Mas desde o período em que foram pela primeira vez propostas, essas teorias acumularam uma grande quantidade de indícios que lhe dão apoio, e quase todos aqueles que têm contato com esses indícios não têm dúvida de que essas teorias são verdades bem estabelecidas.

A indução é racional por definição

Já insisti, contra Popper, que geralmente é racional acreditar nas conclusões das inferências indutivas. No entanto, como eu disse, essa observação não é de modo algum uma solução ao problema da indução. Pois ainda precisamos explicar como as inferências indutivas podem ser racionais dado que suas conclusões não são logicamente garantidas por suas premissas.

Alguns filósofos argumentaram que podemos resolver o problema focando-se no significado comum do termo “racional”.² Afinal, dizem eles, no uso normal, esse termo não se restringe de modo algum ao raciocínio dedutivo. Na verdade, todo mundo reconhece que o raciocínio dedutivo é *uma* espécie de argumento racional. Mas ao mesmo tempo, quase todo mundo também aplica o termo “racional” a outros tipos de raciocínio, e em particular ao raciocínio indutivo.

Para fins ilustrativos, considere três formas diferentes de se prever o tempo. O primeiro tipo de previsor não presta qualquer atenção aos padrões de tempo do passado, e simplesmente estima de maneira aleatória sobre o tempo de amanhã. Um segundo tipo de previsor presta atenção nos padrões do passado, mas prevê que o tempo

2

Veja Paul Edwards, “Russell’s Doubts about Induction”, *Mind*, 68, 141-63; e a seção 9 de P. F. Strawson, *Introduction to Logical Theory* (Londres, 1952).

do futuro baseando-se na suposição de que os padrões de tempo do futuro vão ser diferentes dos padrões do passado: assim, por exemplo, ao ver o céu avermelhado ao entardecer, esse previsor raciocina que, uma vez que o céu avermelhado previu bom tempo no passado, o tempo de amanhã não será bom. O terceiro previsor trabalha sob a suposição de que os padrões de tempo do passado são de fato um guia para padrões futuros, e, portanto, baseando-se na experiência do passado, toma o céu avermelhado no entardecer como um sinal de que haverá bom tempo amanhã.

Agora, se perguntarmos às pessoas que entendem o significado da palavra “racional” qual desses três previsores é racional, não há dúvidas de que responderão que o terceiro previsor é racional, e os outros dois não. E não há dúvidas de que também diriam que, em geral, as pessoas que antecipam o futuro baseando-se no passado são racionais, e aquelas que apenas estimam ou esperam que o futuro seja diferente do passado são irracionais.

Isso não mostra que a indução é racional? Pois o que mais seria necessário para mostrar isso além do fato das pessoas que entendem o significado do termo “racional” concordarem que esse termo é aplicável ao raciocínio indutivo?

Essa forma de argumento é conhecida como “argumento do caso paradigmático”, e foi muito popular entre os filósofos da linguagem comum britânicos nas décadas de 1950 e 1960. Foi aplicado a outros problemas filosóficos além do problema da indução. Assim, por exemplo, em resposta à tese de que os seres humanos na verdade não têm livre-arbítrio, os filósofos da linguagem comum chamaram a atenção que qualquer um que entenda a expressão “agir de acordo com seu livre-arbítrio” não hesitará em aplicá-la a uma série de ações humanas. Afinal, argumentaram os filósofos da linguagem comum, de acordo com o nosso uso comum do termo, ações como beber uma xícara de café ou comprar um carro novo não são casos paradigmáticos de ações livres? E o que mais seria necessário para mostrar que o livre-arbítrio existe além do fato das pessoas que entendem o significado do termo “livre-arbítrio” concordarem que ele se aplica a esse tipo de ações?

Esse exemplo, no entanto, serve também para exibir a fraqueza dos argumentos do caso paradigmático. A única razão de alguns filósofos duvidarem da existência do

livre-arbítrio é porque eles pensam que há uma exigência fundamental para uma ação ser livre, a saber, que não seja determinada por causas do passado e, porque, além do mais, eles duvidam que quaisquer ações humanas não sejam assim determinadas. Qualquer filósofo desse tipo responderá ao argumento do caso paradigmático a favor do livre-arbítrio da seguinte forma: “Talvez as pessoas comuns sejam felizes ao aplicar o termo ‘livre-arbítrio’ a ações como beber uma xícara de café ou comprar um carro novo. Mas isso é apenas porque elas estão assumindo implicitamente que essas ações não sejam determinadas por causas do passado. Mas na verdade elas estão erradas sobre essa suposição. Todas as ações humanas são determinadas por causas do passado. Portanto, não há livre-arbítrio realmente, e as pessoas comuns estão apenas cometendo um erro quando aplicam o termo ‘livre-arbítrio’ como aplicam”.

O mesmo se aplica à tentativa de estabelecer a racionalidade da indução apelando-se ao uso comum. Pois a linguagem comum deixa em aberto a possibilidade de haver alguma exigência fundamental à forma de inferência para considerá-la racional. E as inferências indutivas podem de fato falhar em satisfazer essa exigência, a despeito da inclinação das pessoas comuns aplicarem o “termo” comum à indução.

Uma defesa fiabilista da indução

Perguntemo-nos, então, se há alguma exigência fundamental que uma forma de inferência tem de satisfazer se tiver de ser qualificada como racional. Ora, uma exigência mínima é certamente que as conclusões dessas inferências tenham de ser em geral verdadeiras, se as premissas o forem. O ponto principal das inferências é aumentar o nosso estoque de conhecimento. As inferências produzem conhecimento novo a partir do antigo: tomam o conhecimento antigo como *input* e geram conhecimento novo como *output*. Mas uma forma de inferência falhará nessa tarefa se produzir conclusões falsas, ainda que provida com premissas verdadeiras. Pois em tais casos a inferência não aumentará o nosso estoque de conhecimento, mas antes nos conduzirá ao erro.

É importante reconhecer que essa exigência – a de que as conclusões de uma forma de inferência deveriam, em geral, ser verdadeiras se as premissas o forem – não equivale necessariamente à exigência de que a forma de inferência devesse ser *dedutivamente válida*. Uma forma de inferência é dedutivamente válida se é logicamente

impossível que as conclusões sejam falsas se as premissas forem verdadeiras. Isso é bem mais forte do que a exigência de que, enquanto *questão de fato*, as conclusões nunca sejam falsas quando as premissas forem verdadeiras. Como ilustração, considere esta forma de inferência.

X é um humano

X tem menos de 200 anos

Ela não é dedutivamente válida. É logicamente *possível* que alguém seja um humano e viva por 200 anos. Mas, como é o caso, não há tais seres humanos, e por isso essa forma de inferência nunca nos levará de uma premissa verdadeira para uma conclusão falsa. (Naturalmente, pode-se tornar dedutiva essa forma de inferência adicionando-se a premissas “Todos os humanos vivem menos de 200 anos”. Mas a minha questão é que, ainda que não adicionemos essa premissa, não tornando a inferência dedutiva, ela ainda satisfaz a exigência de nunca partir de premissas verdadeiras para conclusões falsas.)

Usemos o termo “fiável” para a exigência de que premissas verdadeiras devessem sempre produzir conclusões verdadeiras. Deste modo, pode-se pensar nas inferências dedutivamente válidas como sendo inferências que são *necessariamente* fiáveis. Na terminologia dos mundos possíveis, uma inferência dedutivamente válida gerará conclusões verdadeiras a partir de premissas verdadeiras em *todos* os mundos possíveis. Uma inferência fiável embora não-dedutiva, em contraste, sempre gera conclusões verdadeiras a partir de premissas verdadeiras no mundo efetivo, mas poderia não gerar em outros mundos possíveis (digamos, nos mundos em que os humanos vivem mais de 200 anos).

Dado essa distinção, parece claro que a fiabilidade é uma exigência mínima para uma forma de inferência ser aceitável. No entanto, exigir a validade dedutiva parece uma extravagância. Se temos uma forma de inferência que funciona no mundo efetivo, por que exigir também que ela devesse funcionar em qualquer outro mundo possível, ainda que improvável ou remoto?

Essas observações apóiam os filósofos da linguagem comum na insistência deles de que as inferências dedutivamente válidas não esgotam a categoria das inferências racionais. Mas eles oferecem um tipo diferente de razão a fim de reconhecer algumas formas de inferências não-dedutivas como racionais. Os filósofos da linguagem comum estavam dispostos a considerar como racional qualquer forma de inferência que os falantes normais do português chamem de racionais. As observações feitas nesta seção, no entanto, sustentam que uma forma de inferência deveria contar como racional somente se satisfizer as exigências fundamentais de fiabilidade de se transmitir a verdade da premissa à conclusão.

Dever-se-ia dizer que é uma questão de grande controvérsia saber se a fiabilidade é suficiente para a racionalidade. Essa questão é parte de um amplo debate contemporâneo que envolve não apenas a noção de racionalidade, mas também de noções correlatas como *conhecimento* e *justificação*. Penso que poucos filósofos contemporâneos ainda queriam dizer que uma crença é racional (conhecimento, justificada) somente se obtida por meios necessariamente fiáveis (como por inferência dedutiva). Mas entre os remanescentes há uma divisão, entre aqueles (chamemo-los “fiabilistas” daqui para frente) que pensam que uma fonte fiável é por si suficiente para uma que crença seja racional (conhecimento, justificada) e aqueles que pensam que alguma exigência adicional, tal como persuasão intuitiva, também precisa ser satisfeita.

No entanto, não há necessidade de se resolver essa questão mais ampla aqui. De modo que, no restante desta seção vou discutir a seguinte tese condicional: *se* você pensa que a fiabilidade de uma forma de inferência é suficiente para a sua racionalidade, *então* terá uma resposta ao problema da indução.

Note primeiro que, se adotarmos o ponto de vista fiabilista, o argumento original contra a indução deixa de apresentar um problema substancial. Pois o argumento original era simplesmente o de que as premissas de um argumento indutivo não implicam dedutivamente a sua conclusão. Mas uma vez que já não exigimos que os argumentos indutivos sejam logicamente infalíveis, e sim apenas que transmitam a verdade de maneira fiável, esse não é de todo um argumento contra a indução. Pois,

como já enfatizado, uma forma de inferência pode ser fiável sem ser dedutivamente válida.

Essa é apenas parte de uma defesa fiabilista da indução. Pois, ainda que o argumento tradicional falhe em mostrar que a indução não é fiável, o fiabilista ainda precisa fornecer razões para se pensar que a indução é fiável. Ao contrário do filósofo da linguagem comum, o fiabilista não pode simplesmente defender a indução baseando-se no fato de a maioria das pessoas considerá-la “racional”. Pois, de acordo com o fiabilismo, uma forma de inferência é racional somente se satisfaz a exigência fundamental de fiabilidade de transmitir a verdade das premissas à conclusão.

Mas talvez o fiabilista possa responder esse desafio. A questão é saber se as inferências indutivas produzem em geral conclusões verdadeiras caso sejam dadas premissas verdadeiras. O fiabilista pode mostrar que há muitos indícios de que produzem. Quando as pessoas fizeram induções a partir de premissas verdadeiras no passado, argumenta o fiabilista, as suas conclusões se mostraram verdadeiras. Podemos inferir desse indício, portanto, que as inferências indutivas são em geral transmissoras fiáveis de verdade.

É certo que esse último passo é ele mesmo uma inferência indutiva que parte do sucesso das induções no passado à sua fiabilidade geral, e por isso esse argumento é simplesmente uma versão da defesa indutiva da indução que acusei de pressupor o que está em causa na Seção 1.2.2. No entanto, naquele ponto estávamos presumindo que o argumento tradicional levantava um problema genuíno para a indução e que, portanto, seria ilegítimo usar a indução em outras análises filosóficas. Mas a primeira observação feita pela defesa fiabilista da indução, foi que o argumento tradicional, que apenas mostra que a indução não é dedução, em nada compromete a indução. Por que, então, não deveríamos usar os nossos métodos indutivos normais determinar se a indução é fiável? De que outra forma, pode razoavelmente indagar o fiabilista, esperaríamos tratar o problema?

Esse tipo de defesa indutiva da indução claramente não irá persuadir alguém que já não aceita induções, pois tal pessoa não estará propensa a concluir a partir da premissa que as induções funcionaram no passado, que funcionarão no futuro. Mas o

fiabilista pode responder que o argumento indutivo a favor da indução supostamente não cura quaisquer excêntricos que possam rejeitar a indução. Antes, ele supostamente explica, às pessoas normais como nós, como estamos autorizados a pensar que a indução é fiável e, portanto, racional.

Nem todos os filósofos concordariam que essa defesa fiabilista de indução evita a circularidade. Mas a essa altura proponho que deixemos essa questão de lado e nos voltemos, ao invés, a uma objeção mais direta. Essa defesa assume como premissa do argumento indutivo a favor da indução que as induções com premissas verdadeiras têm pelo menos gerado conclusões verdadeiras até agora. Mas isso é mesmo verdadeiro? Não há muitos casos em que as pessoas fizeram induções e, contudo, obtiveram conclusões falsas ao invés de verdadeiras?

Esse é claramente um desafio que os fiabilistas precisam responder. Pois ainda que concedamos a eles a legitimidade dos argumentos indutivos a favor da indução, os fiabilistas não irão a lugar algum se os indícios do passado indicam que a indução não é fiável.

Examinarei dois tipos de razão para se pensar que a indução claramente não é fiável. Um recorre à história da ciência e menciona que muitas teorias científicas indutivamente apoiadas, da astronomia ptolomaica à física newtoniana, mostraram-se falsas devido a indícios posteriores. No entanto, adiarei a discussão desse tipo de argumento histórico contra a indução até a Seção 3. Primeiro, pretendo examinar uma razão mais abstrata para se pensar que a indução, ou pelos a indução enumerativa, não pode ser em geral fiável.

O novo problema da indução de Goodman

Suponha que definamos “verdul” como um termo que se aplica a todos e apenas àqueles objetos que foram *examinados pela primeira vez antes de 2100 e são verdes* ou *que não foram pela primeira vez examinados antes de 2100 e são azuis*.

Imagine agora que queremos averiguar, por meios indutivos, que propriedades, se existir alguma, todas as esmeraldas possuem. Ora, podemos notar que todas as esmeraldas que observamos até agora foram verdes, e concluir com base nisso, através

de uma indução enumerativa, que todas as esmeraldas são verdes. Mas poderíamos também notar que todas as esmeraldas que observamos até agora foram verdus (uma vez que foram examinadas pela primeira vez antes de 2100 e são verdes) e inferir, portanto, através de uma indução enumerativa análoga, que todas as esmeraldas são verdus.

Mas note agora que essas duas conclusões, que todas as esmeraldas são verdes e que todas as esmeraldas são verdus, não podem ser ambas verdadeiras, dado que algumas esmeraldas serão examinadas pela primeira vez apenas depois de 2100. Pois a primeira conclusão implica que essas esmeraldas serão verdes, e a segunda conclusão implica que serão azuis, e por isso uma delas tem de estar errada. Intuitivamente, como é natural, estamos convencidos de que é a “hipótese do verdul” que está errada, e que as esmeraldas ainda serão verdes após 2100. Mas essa suposição intuitiva não é necessária para se alcançar o ponto de partida de Goodman, que é o de que ambas as conclusões foram atingidas através de induções enumerativas da forma: “Um grande número n de As vistos eram Bs”, portanto, “Todos os As são Bs”; contudo, apenas uma dessas conclusões é verdadeira; pelo que nem todas as induções enumerativas podem gerar de modo fiável conclusões verdadeiras.

O verdul é com certeza uma propriedade bastante estranha, e retornarei a isso em breve. Mas a lição central do argumento de Goodman é simplesmente que, a menos que imponhamos algumas restrições sobre quais As e Bs podem entrar nas induções enumerativas, haverá sempre demasiadas induções enumerativas para eles com conclusões verdadeiras. Isso porque, dado qualquer predicado “normal” como “verde” podemos facilmente inventar uma infinidade de predicados estranhos como o verdul que fornecerá conclusões indutivas que têm de ser falsas caso as conclusões indutivas “normais” sejam verdadeiras.

O “novo problema” levantado por Goodman é, então, distinguir, dentre todos aqueles predicados complicados que podem ser definidos, a subclasse que deveria poder entrar nas inferências indutivas. Goodman chamou a isso de o problema de distinguir predicados “projetáveis” de predicados “não-projetáveis”.

Alguns filósofos sugeriram que o problema pode ser resolvido com certa rapidez, proibindo-se simplesmente quaisquer predicados cuja definição faz referência algum tempo particular, como é o caso da definição de “verdul” se referir a 2100. Mas Goodman mostra que o problema não pode ser resolvido com essa facilidade. Pois suponha que definamos “azerde” como “examinado pela primeira vez antes de 2100 e azul ou não examinado pela primeira vez antes de 2100 e verde”. Assim, é verdade que se começarmos com os predicados “verde” e “azul”, e definirmos “verdul” e “azerde” em termos deles, como feito acima, então as definições fazem menção de tempos particulares. Mas suponha que, ao invés, começamos com “verdul” e “azerde” como nossos termos primitivos. “Assim, poderíamos definir “verde” como “examinado pela primeira vez antes de 2100 e verzul ou não examinado pela primeira vez antes de 2100 e azerde”; e poderíamos definir “azul” como “examinado pela primeira vez antes de 2100 e azerde ou não examinado pela primeira vez antes de 2100 e verdul”; dessa perspectiva são, então, as definições de “verde” e “azul” que fazem menção ao tempo. Assim, apelar para o tempo é com efeito pressupor o que está em causa. Pois é apenas porque partimos da suposição de que “verde” e “azul” são predicados consideráveis, no sentido de que “verdul” e “azerde” precisam ser definidos e não vice-versa, que acreditamos que “verdul” e “azerde” não são consideráveis.

A própria solução de Goodman é que os predicados “projetáveis” são simplesmente aqueles cuja ocorrência está “encerrada” em nossas práticas indutivas, no sentido de que eles são os predicados que a comunidade usou para fazer inferências indutivas no passado. Outros filósofos, no entanto, tentaram encontrar meios menos arbitrários de traçar a linha recorrendo às idéias de simplicidade e importância para a ciência. Seria justo dizer, penso, que não há solução universalmente aceita ao problema.

Em qualquer caso, traçar simplesmente uma linha entre predicados projetáveis e os demais é defensavelmente apenas metade do problema. Poderíamos também querer alguma explicação de por que é racional fazer induções com predicados projetáveis e não com os outros. Da perspectiva fiabilista delineada na última seção tal explicação precisaria estabelecer que as induções feitas usando-se predicados projetáveis de modo fiável produzem conclusões verdadeiras dadas premissas verdadeiras.

Há a possibilidade de simplesmente se argumentar mais uma vez que as induções passadas fornecem indícios indutivos a favor da fiabilidade da indução, como foi feito no fim da última seção. Mas já não se pode tomar por garantido que essa manobra seja útil. Pois quando a fizemos na última seção, fizemos via (meta)indução. Mas agora sabemos que a indução enumerativa não é sempre um meio satisfatório de raciocínio e que, na melhor das hipóteses, uma categoria restrita de tais induções é aceitável, designadamente, aquelas que lidam especificamente com as características “projetáveis” do mundo. Até que tenhamos uma teoria mais detalhada da “projetibilidade”, não podemos tomar por garantido que o sucesso das induções passadas seja o tipo de padrão projetável que forneça indícios indutivos a favor de sua própria continuação.

Neste ponto, no entanto, proponho deixar esse tópico. Retornarei e continuarei nele de uma perspectiva algo diferente no fim da seção 3.6.

2. LEIS DA NATUREZA

2.1. Hume, Leis e Acidentes

Nesta seção vou considerar um enigma diferente levantado pela existência de verdades gerais sobre a natureza. O enigma não tem a ver com o nosso conhecimento de tais verdades, mas com a natureza da realidade que elas descrevem: é um problema de metafísica ao invés de epistemologia. Esse problema é normalmente chamado de o problema de distinguir as “leis da natureza” das “generalizações acidentais”.

Um modo útil de abordar esse problema é retornar à análise de David Hume da causalidade. Antes de Hume os filósofos aceitavam que quando uma coisa causa outra, isso era porque a causa possuía algum tipo de poder que necessitava a ocorrência do efeito. Além do mais, eles consideravam que podemos conhecer *a priori* essas conexões de necessitação, no sentido de que podemos inferir *a priori* que o efeito seguirá necessariamente a causa, ainda que nunca tenhamos tido experiência prévia de sua ocorrência.

Hume argumentou contra essa abordagem da causalidade. Ele apontou que quando observamos um evento causando outro (por exemplo, o impacto de uma bola de bilhar causando o movimento da outra), nunca vemos qualquer conexão de necessitação. Tudo que vemos é o evento inicial (o impacto da primeira bola), e então o evento subsequente (o movimento da segunda bola), mas nunca uma terceira coisa que poderia as conectar. Além disso, Hume argumentou que não há conhecimento *a priori* do tipo que tais conexões de necessitação forneceriam. As pessoas que ainda não observaram as bolas de bilhar talvez não possam dizer, numa primeira ocasião em que vêem uma bola em movimento se aproximando de uma bola estacionária, que o impacto fará a bola estacionária se mover ao invés de explodir ou se transformar num duende.

O próprio tratamento que Hume dá à conexão entre a causa e seu efeito é simplesmente que eventos como a causa sempre são seguidos por eventos como o efeito. Da perspectiva de Hume não há qualquer coisa numa sequência particular de causa e efeito além da ocorrência do primeiro evento e da ocorrência do segundo evento após ele. A conexão é simplesmente que essa sequência é uma instância de um padrão geral no qual, para usar a terminologia de Hume, eventos como a causa estão em “conjunção constante” com eventos como o efeito.

Uma consequência da análise de Hume da causalidade é o problema da indução discutido na seção anterior. Antes de Hume presumia-se que poderíamos saber *a priori* que certos resultados sempre se seguiriam de certas causas. De acordo com Hume, no entanto, o conhecimento da causalidade é apenas o conhecimento das conjunções constantes que não resultam de qualquer conexão entre a causa e o efeito. Portanto, o nosso conhecimento da causalidade pode apenas derivar de nossa experiência da causa estando constantemente conjuntada com o efeito. O problema da indução, então, emerge como o problema de que a nossa experiência, que sempre é uma quantidade finita de instâncias causa-efeito passadas, é insuficiente para garantir aquilo que precisamos para o conhecimento causal, designadamente, o conhecimento de que a causa estará constantemente conjuntada com o efeito, não apenas no passado, mas também no futuro.

O problema da indução é um problema acerca do nosso conhecimento de verdades gerais, um problema de epistemologia. Mas a análise de Hume da causalidade também gera um problema sobre a natureza das verdades gerais, um problema de metafísica. O problema é que a análise de Hume da causalidade torna difícil distinguir as leis da natureza genuínas que expressam verdades causais das generalizações acidentais cuja verdade é uma questão de acaso.

De acordo com Hume, uma lei causal é simplesmente uma afirmação da forma “Sempre que A, então B”. No entanto, há verdades dessa forma que não parecem expressar leis. Sempre que vou assistir ao jogo do Arsenal o placar é 0 x 0. Essa é uma afirmação verdadeira da forma “Sempre que A, então B”. E continua verdadeira, pois não tenho ido assistir mais ao Arsenal. Mas claramente não é uma lei causal. Muito embora minha presença tenha sempre sido seguida por zero gols, eu estar em Highbury não impede a capacidade de marcar gols dos jogadores.

Mas por que não? Se tudo o que é exigido para uma lei é que As sejam sempre seguidos por Bs, então por que não é uma lei que não há gols quando vou assistir o Arsenal? Afinal, há, por hipótese, uma perfeita correlação entre a minha presença no Highbury e ninguém marcar.

Esse é o problema de distinguir leis de acidentes. A abordagem humiana da causalidade ameaça a admissão de generalizações acidentalmente verdadeiras na categoria de leis. Precisamos encontrar algum modo de mantê-las afastadas.

Há duas linhas gerais de resposta a esse problema, às quais chamarei “humiana” e “não-humiana”. Os humianos insistem na ideia humiana básica de que as leis causais exprimem conjunções constantes, e não conexões necessárias, e então tentam explicar por que algumas conjunções constantes (as leis) são melhores que outras (os acidentes). Os não-humianos, em contraste, duvidam dessa ideia básica e argumentam a favor de um retorno à perspectiva pré-humana de que a diferença entre as leis e os acidentes é simplesmente de que as leis, mas não os acidentes, exprimem conexões necessárias.

2.2. Condicionais Contrafactuais

No entanto, antes de explorar esses dois tipos de respostas, será útil tratar de uma questão relacionada. Uma diferença frequentemente notada entre leis e acidentes é que as leis, mas não os acidentes, suportam *condicionais contrafactuais*. Uma condicional contrafactual é uma afirmação “se... então...” com a cláusula antecedente falsa. Assim, por exemplo, a afirmação “Se a temperatura tivesse caído a baixo de 0° C, então teria havido gelo sobre a estrada”, feita na ocasião em que a temperatura de fato não caiu a baixo de 0° C e a água não congelou, é uma condicional contrafactual. Na verdade, é uma condicional contrafactual que aceitamos intuitivamente como verdadeira em virtude da lei que a água sempre congela a 0° C. Mas considere agora a condicional contrafactual “Se eu tivesse ido ao jogo do Arsenal, o placar teria sido 0 X 0”, feita sobre uma partida que não fui e que não terminou em empate. Muito embora seja de fato verdadeiro que em todas as ocasiões em que estive presente não houve gols, não aceitamos essa segunda condicional contrafactual como verdadeira pela seguinte razão. Intuitivamente consideramos que a minha presença não teria feito qualquer diferença. Ainda que eu estivesse esta lá, os gols ainda teriam sido marcados.

É esse o sentido em que as leis, mas não os acidentes, suportam contrafactuais. Intuitivamente projetamos leis, mas não acidentes, em situações contrafactuais. No entanto, ao passo que isso é certamente um bom sintoma da diferença entre leis e acidentes, não corresponde a uma explicação da diferença.

A razão é que o significado das contrafactuais é em si uma questão que exige explicação filosófica. Poderíamos começar tal explicação dizendo que as contrafactuais exprimem aquilo que acontece em situações não efetivas. Mas em que sentido situações não efetivas existem? E se não existem, o que torna as afirmações contrafactuais verdadeiras?

Uma teoria filosófica possível das contrafactuais é dizer que as contrafactuais são verdadeiras apenas no caso em que há uma lei conectando a antecedente e a conseqüente. Mas se tomamos essa via no que diz respeito às contrafactuais, então obviamente não podemos continuar a usar as contrafactuais para explicar a diferença lei-acidente. Pois essa teoria das contrafactuais pressupõe essa diferença.

De qualquer forma, há dificuldades conhecidas que uma explicação das contrafactuais em termos de leis enfrenta. Para mencionar apenas uma, considere contrafactuais em que a antecedente é uma negação de uma lei, como por exemplo “Se a força da gravidade fosse inversamente proporcional a r , ao invés de r^2 , então o universo já teria se contraído”. Essa parece uma afirmação contrafactual perfeitamente cogente, mas a noção de uma lei adicional conectando a antecedente e a consequente parece não se aplicar.

Por causa disso, os filósofos contemporâneos têm desenvolvido várias teorias sobre contrafactuais. Uma teoria popular, que se deve a David Lewis,³ recorre à metafísica dos “mundos possíveis”, e diz que a contrafactual “Se A, então B” é verdadeira se e só se o mundo possível “mais próximo” onde A é também um mundo onde B é verdadeiro. Essa é uma teoria atrativa das contrafactuais. Mas se a adotamos, ou qualquer outra teoria similar, então ainda não podemos explicar a diferença lei-acidente em termos de contrafactuais. Pois, uma vez que estamos a explicar as contrafactuais em termos de mundos possíveis, e não de leis, precisaríamos de alguma outra explicação de por que as leis, mas não os acidentes, “se projetam” nos mundos possíveis próximos. Afinal, de acordo com a perspectiva humiana, tanto as leis quanto os acidentes simplesmente exprimem que os As são sempre seguidos por Bs no mundo efetivo. Assim, por que as leis, mas não os acidentes, também nos informam sobre outros mundos não efetivos?

Uma filosofia completa dessas questões combinaria um tratamento da distinção lei-acidente com um tratamento das contrafactuais a fim de produzir uma explicação do porquê as leis mas não os acidentes suportam contrafactuais. Mas até que tenhamos tal tratamento completo, o poder das leis de suportar contrafactuais é parte do problema de explicar a diferença entre leis e acidente, e não a solução.

2.3. Leis como Generalizações Amplas

3

Counterfactuals (Oxford, 1973).

A estratégia humiana, lembre-se, é explicar por que algumas conjunções constantes (leis) são melhores que outras (acidentes). Uma idéia inicial óbvia é que as leis tendem a ser mais gerais do que os acidentes. A verdade de que a água congela a 0° C cobre uma quantidade indefinida, e talvez infinita, de instâncias. Em contraste, a verdade de que nunca há gols quando vou ao jogo da Arsenal se aplica apenas a uma estranha meia dúzia ou mais de casos.

Mas essa não é de fato uma diferença invariável. Pode bem haver leis com apenas poucas instâncias. “Num universo em expansão, a frequência de expansão diminui” presumivelmente tem apenas uma instância, mas não é menos uma lei por causa disso. E é ainda defensável que haja leis que não têm instâncias, como por exemplo, “Um corpo que não está sujeito a quaisquer forças terá aceleração zero”.

Uma idéia relacionada é que os acidentes são excluídos do estatuto de legiforme porque tendem a ser construídos pelo uso de termos que referem indivíduos particulares espaço-temporais, como “David Papineau” e “Arsenal Futebol Clube”, ao invés de termos puramente qualitativos como “água”, “0° C”, e “congela”. Os termos do último tipo se aplicam a quaisquer objetos em qualquer lugar que tenham as propriedades gerais corretas, ao passo que termos não-qualitativos como “David Papineau” são restritos a indivíduos específicos.

Mas isso sequer toca o essencial da questão. Suponha que comecemos como uma generalização accidental verdadeira construída em termos não-qualitativos, como por exemplo, “Quando David Papineau vai ao jogo do Arsenal, o placar é 0 x 0”, e simplesmente substituimos os termos não-qualitativos por descrições qualitativas detalhadas o bastante para apanhar apenas os mesmos indivíduos. Isto é, suponha que substituamos “David Papineau” por “alguém com tal e tal aparência” e “Arsenal Futebol Clube” por “qualquer clube de futebol com tais e tais arquibancadas”, em que os “tais e tais” são longas descrições que me identificam unicamente e identificam unicamente o Arsenal Futebol Clube. Assim, “Sempre que alguém com tal e tal aparência vai ao clube de futebol com tais e tais arquibancadas” seria uma generalização verdadeira construída em termos puramente qualitativos. Mas ainda seria um acidente.

2.4. As Leis são Indutivamente Apoiadas por suas Instâncias

Não haveria, contudo, a despeito dos argumentos da seção anterior, algum sentido em que os acidentes são demasiado específicos e locais para funcionar como guias gerais ao funcionamento do universo? J. L. Mackie argumentou a favor de um modo diferente de capturar essa intuição. O problema com os acidentes, de acordo com Mackie, não é que eles têm muito poucas instâncias, como tal, mas antes, é que não são *indutivamente apoiados* por suas instâncias. Quando observamos diversos casos de a água congelando a 0° C, isso nos dá boas razões para supor que toda água congela a 0° C. Em contraste, que os times não marcaram nas três primeiras das quatro vezes em que fui ao jogo do Arsenal parece uma razão ruim para supor que a minha presença os impediria de marcar da próxima vez que eu fosse.

Com efeito, Mackie está sugerindo que expliquemos a diferença entre leis e acidentes em termos da diferença entre predicados projetáveis e não-projetáveis.⁴ Recorde-se da discussão do “novo problema da indução” de Goodman na Seção 1. Goodman mostra que precisamos reconhecer a distinção entre padrões que envolvem predicados como “... é verde”, que podem ser racionalmente projetados em outros casos inobservados, e padrões que envolvem predicados como “... é verdul”, que é irracional esperar continuar. A sugestão de Mackie, então, é simplesmente que as leis são aquelas generalizações que contêm predicados projetáveis.

Note como essa sugestão produz uma explicação natural do porquê os exemplos de acidentes tendem a ser construídos em termos não-qualitativos e a ter uma quantidade finita de instâncias. De acordo com Mackie, enquanto que as leis podem ser asseridas com base nos subconjuntos de suas instâncias, os acidentes, que não são indutivamente apoiadas por suas instancias, podem apenas ser aceitos como verdadeiros quando sabemos que checamos exaustivamente todas as suas instâncias. (Por exemplo, apenas soubemos que a generalização do Arsenal foi verdadeira porque eu poderia lhe prometer que não iria mais aos jogos.)

Portanto, uma condição para saber que um acidente é verdadeiro é que ele tenha uma quantidade finita de instâncias, pois do contrário o exame exaustivo seria

4

J. L. Mackie, *Truth, Probability and Paradox* (Oxford, 1973).

impossível. E um modo natural de assegurar tal finitude é construir exemplos de acidentes em termos não-qualitativos. (Não é necessário excluir os acidentes verdadeiros com uma quantidade infinita de instâncias. O ponto é apenas que tais acidentes não podem ser conhecidos e, portanto, não estarão disponíveis como exemplos para discussão filosófica.)

Podemos agora ver exatamente por que os acidentes são inúteis como guias para o funcionamento do universo. Não é que os acidentes sejam menos verdadeiros do que as leis, e nem que sejam necessariamente menos gerais. O que ocorre é apenas que nunca estamos em posição de *usá-los* como guias, pois nunca estamos em posição de confiar numa generalização verdadeira até já termos averiguado tudo o que ela poderia nos dizer por meios independentes.

2.5. Leis e Sistematização

Darei atenção agora a uma abordagem humiana diferente da diferença lei-acidente. No fim desta seção vou compará-la com a abordagem de Mackie. A idéia central é que as leis, mas não os acidentes, são parte duma abordagem científica dos modos como o mundo funciona: a diferença entre “A água congela a 0° C” e “Não há gols quando David Papineau vai ao Highbury” é que o primeiro, mas não o último, é explicável em termos de princípios científicos básicos.

Essa sugestão naturalmente precisa fornecer um tratamento independente dos “princípios científicos básicos” não os considerando com *leis* básicas. Isso é feito apelando-se à idéia de *sistematização mais simples* de verdade gerais. Imagine que de um ponto de vista de Deus, por assim dizer, haja uma classe de generalizações objetivamente verdadeiras que incluem todas as leis e todos os acidentes. Pense agora nos vários modos como essas verdades poderiam ser organizadas num sistema dedutivo baseado num conjunto de axiomas. Algumas dessas sistematizações teriam um grau maior de simplicidade do que outras. (Podemos considerar que quanto menos axiomas, mais simples o sistema.) Mas a simplicidade pagaria o preço por deixar algumas

generalizações fora da sistematização. (Poderíamos incluir *todas* as verdades gerais no sistema simplesmente tomando-as como axiomas. Mas essa sistematização careceria completamente de simplicidade.) Haverá, defensavelmente, uma sistematização que combina favoravelmente força e simplicidade, e que tem um pequeno número de axiomas, por simplicidade, e não obstante trabalha para incluir quase toda a classe de verdades gerais como teoremas que se seguem desses axiomas. Podemos então distinguir leis de acidentes dizendo que os axiomas e teoremas nessa sistematização favorável são leis, ao passo que as verdades gerais deixadas de lado são os acidentes.

Em suma, dizemos que as leis são aquelas verdades gerais que se seguem dos axiomas da ciência, e então usamos o argumento da “simplicidade mais a força” para identificar esses axiomas.

Como essa idéia, que foi primeiramente desenvolvida por F. P. Ramsey no início do século XX, e mais tarde restabelecida por David Lewis, se relaciona com a sugestão de Mackie? Aceitemos, para fins de comparação, que a classe dos predicados projetáveis coincide com a classe que aparece na sistematização “o mais simples mais o mais forte”. Ainda que façamos essa suposição, a teoria de Mackie difere da de Ramsey e da de Lewis. Pois Mackie diz que *qualquer* generalização verdadeira construída com predicados projetáveis é uma lei; ao passo que Ramsey e Lewis exigem além disso que a generalização seja dedutível dos axiomas da ciência.⁵ Assim, para decidir entre essas duas teorias da legiformidade, precisamos considerar o estatuto de alguma generalização que seja construída em termos projetáveis, mas que não seja de fato dedutível dos axiomas da ciência.

Por exemplo, imagine que você esteja fazendo alguma pesquisa com algum equipamento eletrônico complicado, e nota que, quando o equipamento e seu rádio estão ligados ao mesmo tempo, o rádio faz um barulho estranho. Suponha também que essa é a única vez que esse tipo de equipamento complicado será montado, pois você o desmontará no fim do experimento. Dado que as propriedades dos equipamentos

5

F. P. Ramsey, “Universals of Law and Universals of Fact” (1928); reimpresso em *Foundations*, ed. D. H. Mellor (London, 1978); David Lewis, *Counterfactuals* (Oxford, 1973).

eletrônicos e dos rádios são presumivelmente projetáveis, se algo o for, você infere que, quando um equipamento desse tipo estiver ligado, os rádios como o seu farão um barulho estranho. Mas suponha que na verdade não há conexão real, e que o seu rádio esteja fazendo os barulhos estranhos por uma razão completamente diferente. Então, a generalização “Quando um equipamento desse tipo está ligado, os rádios como o seu fazem um barulho estranho” será verdadeira sem exceção, e conterà predicados projetáveis. Contudo, ela claramente não é uma lei. Isso mostra que Ramsey e Lewis estão corretos sobre as leis e Mackie errado, uma vez que a teoria de Ramsey-Lewis não considera essa generalização como uma lei, ao passo que a teoria de Mackie considera. (Se você fosse o experimentador no exemplo, não duvidaria que o padrão é uma lei, pois não duvidaria de que tem uma explicação em termos de ciência básica. Mas contudo, você estará errado ao pensar isso, uma vez que não há tal explicação.)

2.6. A Alternativa Não-Humiana

Uma objeção à teoria das leis de Ramsey-Lewis é que a sua dependência das noções de “força” e “simplicidade” a torna vaga e subjetiva. Mas ainda que deixemos isso passar, e concedamos que a teoria produza um modo razoavelmente preciso de distinguir as generalizações verdadeiras que se qualificam como leis, há outra objeção, uma objeção que de fato pode ser levantada a todas as teorias humianas. A saber, de que toda a abordagem humiana à legiformidade é altamente contra-intuitiva.

Considere estas duas seqüências: (1) A temperatura cai a baixo de 0° C, e então a água congela; (2) Vou ao Highbury, e então não há gols. Os humianos dizem que a única distinção entre elas é que ao passo que são ambas instâncias de generalizações universais verdadeiras, a generalização que cobre (1) é de algum modo mais significativa do que a que cobre (2). Mas isso certamente vai contra a intuição. Pois parece deixar de lado a idéia de que em (1) o primeiro evento faz o segundo acontecer; ao passo que em (2) não há tal conexão entre os dois eventos. Dizer que essa diferença é uma diferença nas generalizações cobertas parece colocar a diferença no lugar errado, tornando-a uma questão linguística ao invés de um aspecto da natureza. Intuitivamente, a questão é saber se há uma conexão na natureza entre os eventos particulares, e não se as

generalizações cobertas são suficientemente gerais, ou indutivamente apoiadas por suas instâncias, ou mesmo se são parte de sistematização favorável.

Favorecer a intuição aqui é simplesmente rejeitar a análise de Hume da causalidade. Mas diversos filósofos contemporâneos têm argumentado que deveríamos fazer justamente isso. Nas últimas décadas David Armstrong, Fred Dretske e Michael Tooley⁶ argumentaram que as leis causais não são simplesmente enunciados de conjunção constante, mas antes exprimem *relações de necessitação* entre as propriedades envolvidas. Eles dizem que o modo de representar o conteúdo de uma lei causal não é simplesmente por “Todos os As são (como ocorrem) seguidos por Bs”, mas antes por “Nec (A, B)” em que “Nec” representa a relação de necessitação entre as propriedades A e B. Assim, no par contrastante acima, a baixa temperatura necessita o congelamento, mas a minha presença no Highbury não necessita a falta de gols.

De acordo com a perspectiva de Armstrong-Dretske-Tooley, uma relação de necessitação entre A e B implica certamente que todos os As são Bs. Mas a implicação conversada não se sustenta: pode haver casos em que todos os As são Bs muito embora não seja verdade que Nec (A, B) – a saber, quando é um acidente que todos os As são Bs.

Assim, essa perspectiva não-humiana oferece uma explicação inteiramente direta da diferença lei-acidente. A diferença é simplesmente a de que as leis exprimem algo que as generalizações acidentalmente verdadeiras não exprimem, a saber, a existência de uma relação de necessitação entre propriedades.

Dada a possibilidade dessa solução simples, a pergunta óbvia a se fazer é por que a maioria dos filósofos nos 250 anos desde Hume não se valeram dela.

Hume tinha dois argumentos contra a idéia de que as leis causais envolvem conexões de necessitação. Primeiro, nunca vemos tais conexões. Segundo, não podemos conhecer as leis da natureza *a priori* como seria possível caso esprimissem necessidades.

6

David Armstrong, *What is a Law of Nature?* (Cambridge, 1983); Fred Dretske, “Laws of Nature”, *Philosophy of Science*, 44 (1977), 248-68; Michael Tooley, “The Nature of Laws”, *Canadian Journal of Philosophy*, 7 (1977), 667-98.

Não precisamos nos demorar no primeiro argumento de Hume. A suposição de que não podemos falar algo com sentido sobre coisas que não podemos observar tem tido poucos defensores neste século, ainda que fosse geralmente aceita na época de Hume. O exemplo da ciência contemporânea, com sua conversa sobre átomos, elétrons, e ondas de rádio tem mostrado que a referência dotada de significado não se restringe aos fenômenos observáveis. Portanto, o fato de que não podemos ver as conexões de necessitação não significa automaticamente que não possamos falar sobre elas.

O segundo argumento merece mais atenção. Esse argumento presume que se as leis exprimem necessidades, então têm de ser conhecíveis *a priori* (e por isso conclui que, uma vez eu as leis claramente não podem ser conhecidas *a priori*, não podem exprimir necessidades). A suposição de que a necessidade implica a aprioricidade permaneceu incontestada até muito recentemente na tradição filosófica ocidental. No início da década de 1970, no entanto, o filósofo americano Saul Kripke argumentou que a noção *metafísica* de necessidade precisa ser claramente separada da noção *epistemológica* de aprioricidade. Em particular, Kripke argumentou que muitas afirmações de necessidade (por exemplo, “A Estrela da Tarde = a Estrela da Manhã”) são necessárias (pois como *poderia* esse planeta não se ele mesmo?), muito embora só após descobertas empíricas *a posteriori* é que se pode saber que são verdadeiras.

É surpreendente Armstrong, Dretske e Tooley terem desenvolvido essa perspectiva não-humiana das leis num período de cinco anos após a publicação das idéias de Kripke. Isso sugere que a chave que os permitiu rejeitar a perspectiva de Hume das leis estava na separação da necessidade da aprioricidade. Pois quando dizem que as leis da natureza exprimem que A necessita B, certamente não querem dizer que essas leis possam ser conhecidas *a priori*. Nesse sentido, o tipo de conexão necessária que eles defendem é diferente do tipo que Hume rejeitou. (Significa também que a perspectiva deles das leis não faz diferença para o problema da indução: uma vez que as leis têm de ser derivadas de indícios *a posteriori*, ainda precisamos explicar como os indícios passados podem nos informar algo que implica padrões futuros.)

Embora pareça altamente plausível que as perspectivas de Kripke acerca da necessidade tenham motivado o ressurgimento das perspectivas não-humianas das leis

da natureza, há diferenças importantes entre esses dois desenvolvimentos. Mais especificamente, as conexões de necessitação não-humianas não são na verdade *necessárias* no sentido de Kripke. As necessidades kripkianas supostamente são obtidas em todos os mundos possíveis. É simplesmente impossível que um planeta existisse sem ser ele mesmo. Mas os não-humianos contemporâneos não exigem que as leis da natureza sejam necessárias nesse sentido. Eles concedem que seja possível que a força da gravidade pudesse ter sido mais fraca do que é, que a água pudesse congelar numa temperatura diferente, e assim por diante. A idéia deles de uma conexão de necessitação é a de que uma propriedade faz outra acontecer, e não a ideia kripkiana de uma afirmação que *não poderia ser possivelmente falsa*.

Essa diferença aponta para uma dificuldade que as perspectivas não-humianas das leis da natureza enfrentam. Os não-humianos dizem que a necessitação envolve algo além da conjunção constante: se dois eventos estão relacionados por necessitação, então segue-se que estão constantemente conjuntados; mas dois eventos podem estar constantemente conjuntados sem estarem relacionados por necessitação, como quando a conjunção constante é uma questão de acidente. Portanto, a necessitação é uma relação mais forte que a conjunção constante. No entanto, os não-humianos dizem muito pouco sobre essa força extra. Dizem-nos que não é a necessidade no sentido kripkiano de verdade em todos os mundos possíveis. Mas não nos dão qualquer caracterização positiva dessa força extra, exceto que distingue leis de acidentes. Os críticos da perspectiva não-humana argumentam que uma abordagem satisfatória das leis deve lançar mais luz sobre a natureza das leis do que tal perspectiva. Eles se queixam que a noção de necessitação simplesmente reitera o problema ao invés de resolvê-lo.

Portanto, podemos resumir a nossa discussão geral das leis da natureza numa escolha. Se você quer explicações, e não faz caso das intuições, então pode recorrer a uma estratégia humiana, com a teoria de Ramsey-Lewis como a versão mais promissora. Mas se você quer uma abordagem das leis da natureza que se adéqüe às nossas intuições pré-teóricas, e faz pouco caso da queixa de que ela simplesmente reitera a diferença lei-acidente sem explicá-la, então pode adotar a opção não-humana contemporânea.

3. REALISMO, INSTRUMENTALISMO E SUBDETERMINAÇÃO

3.1. Instrumentalismo versus Realismo

Na primeira seção discuti o problema da indução. Nesta, pretendo considerar uma dificuldade diferente para o nosso conhecimento do mundo natural e em particular do conhecimento científico. Grande parte da ciência consiste de afirmações sobre entidades inobserváveis como vírus, ondas de rádio, elétrons e quarks. Mas se essas entidades são inobserváveis, como os cientistas supostamente as descobriram? Se eles não podem vê-las ou tocá-las, não se segue disso que as suas afirmações sobre elas são na melhor das hipóteses conjecturas especulativas ao invés de conhecimento sólido?

É importante distinguir o problema da inobservabilidade do problema da indução. Ambos podem ser vistos como dificuldades ao conhecimento teórico na ciência. Mas o problema da indução surge porque as teorias científicas fazem afirmações gerais, enquanto que o problema da inobservabilidade se deve à nossa falta de acesso sensível ao conteúdo de muitas teorias científicas. (Assim, o problema da indução surge de afirmações gerais ainda que não sejam sobre inobserváveis como “Todo sódio produz uma chama alaranjada”. Conversamente, o problema da inobservabilidade surge de afirmações sobre inobserváveis ainda que não sejam gerais, tal como “Há um elétron livre nesta gota de óleo”. Nesta seção e na próxima, no entanto, será conveniente usar o termo “teoria” especificamente para afirmações sobre inobserváveis ao invés de usá-la para afirmações gerais de qualquer tipo.)

Há duas escolas de pensamento que tratam do problema da inobservabilidade. De um lado estão os *realistas*, que pensam que o problema pode ser resolvido. Os realistas argumentam que os fatos observáveis fornecem bons indícios indiretos da existência de entidades inobserváveis, e concluem, portanto, que as teorias científicas podem ser consideradas como descrições corretas do mundo inobservável. Do outro estão os *instrumentalistas*, que sustentam que não estamos em posição de fazer juízos sólidos sobre mecanismos imperceptíveis. Os instrumentalistas aceitam que as teorias sobre esses mecanismos possam ser “instrumentos” úteis para simplificar os nossos cálculos e produzir previsões. Mas eles argumentam que essas teorias não são descrições mais verdadeiras do mundo do que a “teoria” de que toda a matéria numa pedra está

concentrada em seu centro de massa (que é também uma suposição extremamente útil para se fazer certos cálculos, mas que é claramente falsa).

No início no século XX os instrumentalistas costumavam argumentar que nem mesmo deveríamos interpretar literalmente as afirmações teóricas pela razão de que não podemos sequer falar algo com sentido sobre entidades que nunca observamos diretamente. Mas como eu disse na última seção, o desenvolvimento da ciência contemporânea, com a sua conversa sobre átomos, elétrons, e assim por diante, tornou essa restrição sobre a conversa dotada de sentido difícil de ser defendida. Por isso, esse tipo de instrumentalismo *semântico* está fora de moda hoje. Os instrumentalistas contemporâneos aceitam que os cientistas possam postular de modo significativo, digamos, que a matéria é constituída de minúsculos átomos contendo núcleos orbitados por elétrons. Mas assumem uma atitude cética a tais postulados, dizendo que não estamos autorizados a acreditar neles (opondo-se ao seu uso com um instrumento para cálculos.)

3.2. Primeiros Argumentos a favor do Realismo

Uma linha inicial de argumento aberta ao realismo é identificar algumas características da prática científica e então argumentar que o instrumentalismo é incapaz de explicá-las. Assim, por exemplo, os realistas têm apontado para o fato de que os cientistas caracteristicamente procuram unificar diferentes tipos de teoria na busca de uma única “teoria de tudo”. No século XIX, por exemplo, os físicos que trabalhavam com a termodinâmica desenvolveram a teoria cinética dos gases, que explicou as variações na temperatura, na pressão, e no volume dos gases postulando que os gases são constituídos de aglomerados de minúsculas partículas; ao mesmo tempo, os químicos estavam desenvolvendo a teoria atômica da matéria, que explicava as combinações químicas baseando-se na suposição de que a matéria era feita de átomos, um tipo de átomo para cada elemento. Uma questão óbvia era investigar a relação entre as duas teorias: eram as partículas dos físicos combinações de átomos, e se eram, quais os tipos de combinação? A solução para essa questão nem sempre foi fácil, mas com o tempo uma conclusão satisfatória foi alcançada.

Todo esse processo, no entanto, só faz sentido com base na suposição de que as teorias são descrições *verdadeiras* da realidade, indica o realista. Afinal, diz o realista, se as teorias são simplesmente máquinas de calcular úteis, então por que esperar que diferentes teorias sejam unificadas num relato consistente? A unificação é claramente desejável caso todas as nossas teorias visem contribuir para a verdade geral, porém não parece haver razão semelhante pela qual uma penca de instrumentos devesse ser unificável num grande “instrumento de tudo”.

Outras características da ciência a que os realistas recorrem como argumentos contra o instrumentalismo incluem o uso das teorias para explicar fenômenos inobserváveis e a confiança nas teorias ao se fazer novas previsões. Considerarei isso em sequência. O tópico sobre explicação será discutido em detalhe na seção 5. Mas por agora precisamos apenas notar que os cientistas frequentemente explicam o comportamento dos fenômenos observáveis em termos de mecanismos inobserváveis. Assim, para usar um dos exemplos acima, os cientistas explicam porque a pressão de um gás ideal aumenta quando sua temperatura aumenta referindo-se ao comportamento de minúsculas partículas que compõem o gás. Mas certamente que isso só faz sentido se as minúsculas partículas realmente existirem e a teoria que as descreve não for apenas uma instrumento para se fazer cálculos, sublinha o realista. Certamente que não podemos dizer que a pressão aumenta porque as partículas minúsculas estão se movendo rapidamente, se não acreditamos na existência dessas partículas.

Eis, então, o argumento da previsão. Os cientistas, com base em suas teorias, prevêem frequentemente fenômenos observáveis surpreendentes e até então completamente desconhecidos. Por exemplo, Einstein previu, com base na teoria geral da relatividade, que a luz se curvaria nas proximidades do sol. Fora essa teoria, não havia quaisquer razões para se esperar isso. Contudo, essa previsão foi triunfantemente confirmada pelas famosas observações de Sir Arthur Eddington no oeste da África durante um eclipse solar em 1919. Isso fornece outro argumento a favor do realismo. Pois o realista pode insistir que não haveria razão pela qual as previsões devessem sequer funcionar se as teorias por trás delas não fossem verdadeiras.

Esses três argumentos, o da unificação, o da explicação e o da previsão, dão algum apoio ao realismo. Mas nenhum deles é conclusivo. Em cada caso há duas possíveis linhas de resposta disponível aos instrumentalistas. Eles podem oferecer uma explicação instrumentalista da característica relevante da prática científica. Alternativamente, eles podem negar que essa característica seja de fato parte da prática científica em primeiro lugar. Passarei por esses três casos em sequência.

3.3 Primeiras Respostas Instrumentalistas

3.3.1. Unificação

Primeiro o argumento da unificação. A primeira possibilidade para os instrumentalistas é oferecer uma explicação instrumentalista da prática científica de teorias unificadoras. Eles podem fazer isso argumentando que a unificação na ciência é motivada, não pela busca de uma verdade subjacente, mas simplesmente pelo desejo de um instrumento único e com o propósito de calcular, ao invés de uma mixórdia de diferentes instrumentos para diferentes problemas. Se o objetivo das teorias é a utilidade ao invés da veracidade, não é mais útil ter um dispositivo que lidará com todos os problemas ao invés de ter de se preocupar com qual ferramenta será mais adequada para o problema em questão?

A segunda possibilidade para um instrumentalista frente ao argumento da unificação é negar que a unificação seja essencial à ciência. Assim, em *How the Laws of Physics Lie*,⁷ Nancy Cartwright argumenta que a ciência é na verdade uma mixórdia de diferentes instrumentos. Ela sustenta que diante de um dado problema os cientistas comumente lançam mão de técnicas simplificadoras e princípios básicos que nada têm

7

(Oxford, 1983).

de teoria geral, mas que os mostra a resposta correta ao tipo de problema em questão. Na perspectiva de Cartwright, portanto, a unificação não é central à ciência em primeiro lugar, e por isso os instrumentalistas não precisam fazer qualquer coisa para explicá-la.

3.3.2. Explicação

As mesmas duas linhas de resposta podem ser oferecidas ao argumento realista da explicação. Aqui, a linha de resposta mais normal é a segunda, a saber, negar que a explicação seja de fato uma característica essencial da prática científica. Os instrumentalistas podem argumentar que o objetivo essencial da ciência é descrever, e não explicar. Aquilo que queremos da ciência, dirão, é um relato acurado de como o mundo observável se comporta. A questão adicional do porquê ele se comporta como comporta é uma questão mais difícil que nos deixa aquém da ciência caso possa ser respondida de todo. (Afinal, o instrumentalista pode alegar, mesmo os realistas têm de parar para explicar o mesmo ponto. Talvez eles possam explicar os observáveis em termos de inobserváveis, e alguns inobserváveis em termos de outro. Mas mesmo os realistas terão de admitir que em algum ponto, talvez com quarks ou outras partículas fundamentais, eles esgotam a explicação e podem apenas descrever o comportamento das partículas fundamentais sem ter de explicá-las em termos de mecanismos adicionais.)

Como eu disse, esse tipo de negação de que a explicação seja essencial à teorização científica é a resposta instrumentalista normal ao argumento da explicação. Mas uma minoria de instrumentalistas tenta a direção oposta, e argumenta que não há qualquer coisa na explicação que o instrumentalismo não possa dar conta. De acordo com os instrumentalistas dessa estirpe, é um erro pensar na explicação científica com uma questão de identificar as causas genuínas por trás dos fenômenos observáveis como oposta a mostrar simplesmente como esses fenômenos são parte de um padrão mais amplo. O cientista que “explica” as variações na pressão dos gases pela teoria cinética não está, dessa perspectiva, especificando as verdadeiras causas inobserváveis dessas variações, mas simplesmente mostrando como elas se conformam às mesmas equações subjacentes assim como outros tipos de comportamento observável de gases. (Talvez essa segunda resposta ao argumento da explicação faça pouco mais do que inventar um

novo significado para “explicação”. Mas se isso o deixa preocupado, sempre há a primeira resposta para se recorrer.)

3.3.3. *Previsão*

Resta o argumento realista da previsão. Aqui duas linhas instrumentalistas de resposta estão disponíveis novamente. A mais radical, e talvez menos plausível, seria negar que a capacidade de fazer tais previsões seja uma característica genuína da prática científica. Os instrumentalistas que seguem essa linha com certeza consentem que os cientistas façam “previsões” no sentido de que tirem conseqüências observacionais de suas teorias. Mas podem negar que essa prática gere quaisquer previsões mais *verdadeiras* do que suposições aleatórias gerariam. Afinal, eles podem salientar, as únicas previsões verdadeiras das quais lembramos são as bem sucedidas, como a previsão de Einstein do desvio da luz. Mas para toda previsão bem sucedida há milhares de experimentos científicos que não produzem os resultados esperados. Assim, que razões reais temos para pensar que as teorias sobre inobserváveis nos permitam antecipar novos fenômenos observáveis? Talvez essa seja apenas uma impressão criada pela memória seletiva. Se isso estiver correto, e a ciência não for de fato preditivamente bem sucedida, então obviamente não há necessidade de uma explicação instrumentalista desse sucesso.

No entanto, como eu disse, essa resposta não é inteiramente plausível. Parece improvável que a capacidade das teorias sobre inobserváveis às vezes antecipar novos fenômenos observáveis seja apenas uma questão de sorte. No entanto, ainda que aceitemos que a ciência seja preditivamente bem sucedida, resta espaço para uma abordagem instrumentalista disso. A abordagem realista, lembre-se, foi a de que as teorias sobre inobserváveis são caracteristicamente verdadeiras, e por isso não surpreende que resultem em previsões verdadeiras. Os instrumentalistas, que negam a verdade das teorias sobre inobserváveis, não podem dizer isso. Mas podem dizer algo mais. Eles podem aceitar que haja um padrão bem estabelecido, visível na história da ciência, de novas previsões observáveis sugeridas pelas teorias sobre inobserváveis que se mostraram verdadeiras. E então podem simplesmente insistir, de acordo com o seu instrumentalismo, que não há necessidade de fornecer qualquer explicação ulterior

desse padrão em termos de tais fatos subjacentes como a verdade das teorias tratadas. Afinal, o instrumentalismo é precisamente o ponto de vista de que não precisamos explicar os padrões manifestos em termos de causas subjacentes (ou que no máximo deveríamos “explicá-las” conectando-as a padrões manifestos mais amplos). Dado que os instrumentalistas partem da negação da necessidade das explicações inobserváveis, seria cometer uma petição de princípio contra eles insistir que eles deveriam produzir tal explicação do sucesso preditivo da ciência.

3.4. A Subdeterminação da Teoria pelos Dados

Na última seção argumentei que podemos resistir a vários argumentos contra o instrumentalismo. Vou agora deixar que o instrumentalismo continue na ofensiva e considerar alguns argumentos positivos contra o realismo. Há duas fortes linhas de argumento que os instrumentalistas podem usar para lançar dúvida sobre o realismo. Nesta seção e na próxima discutirei “a subdeterminação da teoria pelos indícios” e algumas questões correlatas. Na Seção 3.6 considerarei “a metaindução pessimista da falsidade passada”. Na verdade, penso que nenhum desses argumentos sejam bem sucedidos em tornar o realismo duvidoso. Mas são argumentos que merecem meticulosa consideração.

O argumento da subdeterminação vindica que, dada qualquer teoria sobre inobserváveis que se adéqua aos fatos observáveis, haverá outras teorias incompatíveis que se adéquam aos mesmos fatos. E por isso, conclui o argumento, nunca estamos em posição de saber que qual dessas teorias é a verdadeira.

Por que deveríamos aceitar que há sempre mais que uma teoria que se adéqua a qualquer conjunto de fatos observáveis? Há duas rotas para essa conclusão. Uma vem da tese de Duhem-Quine, originalmente formulada pelo filósofo e historiador francês Pierre Duhem na virada do século XIX e mais tarde reavivada pelo lógico americano W. V.

Quine.⁸ Duhem e Quine apontaram que uma teoria científica T (como a teoria newtoniana da gravitação) normalmente não implica previsões P por si própria (sobre os movimentos dos planetas, digamos), mas apenas em conjunção com hipóteses auxiliares H (envolvendo coisas como o número dos outros planetas, as suas massas, a massa do sol, e assim por diante).

$$(T \& H) \rightarrow P$$

Por causa disso, T sempre pode ser defendida frente a observações contrárias (como a conhecida anomalia da teoria newtoniana apresentada pela órbita de Mercúrio) ajustando-se as hipóteses auxiliares H (postulando um planeta até então inobservado, digamos, ou uma distribuição não homogênea de massa no sol). O ponto é que a refutação observacional de P não refuta T, mas apenas a conjunção T & H.

$$\text{Não-P} \rightarrow \text{não-(T \& H)}$$

Assim, T pode ser preservada, e de fato ainda explicar não-P, desde que substituamos H por alguma hipótese alternativa H', tal que

$$(T \& H') \rightarrow \text{não-P.}$$

Isso produz a tese de Duhem-Quine: Qualquer afirmação teórica T pode ser preservada consistentemente frente a um indício contrário fazendo-se ajustes em algum lugar do nosso sistema de crenças. A subdeterminação da teoria pelos indícios (SDTI) segue-se prontamente. Pois a tese de Duhem-Quine parece implicar que os partidários de teorias rivais sempre conseguirão manter suas respectivas posições frente a qualquer dado observacional. Imagine duas teorias rivais T₁ e T₂. Por mais que os indícios se acumulem, as versões de T₁ e T₂ conjuntadas, quando necessário, a hipóteses auxiliares amplamente revistas permanecerão consistentes com os dados embora inconsistentes entre si.

8

P. Duhem, *The Aim and Structure of Physical Theory*, edição em inglês (London, 1962)/edição brasileira: *A Teoria Física: seu objeto e sua estrutura*. Eduerj, 2014; W. V. O. Quine, "Two Dogmas of Empiricism", in *From a Logical Point of View* (Cambridge, Mass., 1953)/edição brasileira: *De um Ponto de Vista Lógico*. Unesp, 2011.

A outra rota para a SDTI, primeiramente desenvolvida por físicos como Henri Poincaré na virada do século XIX, tem um ponto de partida diferente.⁹ Começa, não com duas teorias rivais, mas com uma dada teoria, cujas previsões observacionais são supostamente exatas. Imagine que T_1 é a verdade completa acerca da realidade física e que implica verdades observacionais O . Então, podemos sempre construir uma T_2 “desocamizada” que postule mecanismos inobserváveis mais complicados, mas que, não obstante, tem precisamente as mesmas conseqüências observacionais.

Por exemplo, suponha que comecemos com suposições padrões sobre a localização dos corpos no espaço-tempo e sobre as forças que agem sobre eles. Uma teoria desocamizada poderia então postular que todos os corpos, incluindo todos os instrumentos de medição, estão se acelerando a $0,3 \text{ m/s}^2$ numa dada direção, e então adicionar apenas as forças extras necessárias para explicar isso. Essa teoria teria clara e exatamente as mesmas conseqüências observacionais que a teoria original, muito embora a contradiga no nível inobservável.

Para ver a diferença entre os dois argumentos a favor da SDTI, note que o argumento de Duhem-Quine não especifica exatamente com que teorias completas terminaremos, uma vez que deixa em aberto como as hipóteses auxiliares de T_1 e T_2 podem ser revistas; o argumento da desocamização, em contraste, especifica de fato T_1 e T_2 de maneira detalhada, incluindo as hipóteses auxiliares. Em compensação, o argumento de Duhem-Quine nos promete teorias alternativas qualquer que seja o dado observacional que possa aparecer no futuro; ao passo que o argumento da desocamização presume que todas as observações futuras são como T_1 prevê.

3.5. Simplicidade e Eliminação

A minha opinião é a de que os argumentos da seção anterior nos dão boas razões para aceitar a SDTI, a tese de que sempre haverá teorias incompatíveis pra explicar qualquer corpo de fatos observacionais. Não concordo, no entanto, que a SDTI seja um bom argumento contra o realismo. O que a SDTI mostra é que mais do que uma teoria

9

H. Poincaré, *Science and Hypothesis*, edição em inglês (New York, 1952). Edição brasileira: *A ciência e a Hipótese*. Editora UnB, 1988.

sobre inobserváveis sempre se adequará a qualquer conjunto de dados observacionais. Mas é demasiado apressado concluir, como muitos filósofos fizeram, que isso torna o realismo sobre inobserváveis insustentável. Pois, deveríamos reconhecer que não há qualquer coisa nos argumentos a favor das teorias subdeterminadas alternativas que mostre que essas teorias alternativas serão sempre igualmente bem apoiadas pelos dados. O que os argumentos mostram é que as diferentes teorias sempre serão consistentes com os dados. Mas não excluem a possibilidade de que dentre essas teorias alternativas uma seja amplamente mais plausível que as outras e que, por essa razão, deveríamos acreditar que ela seja verdadeira. Afinal, os defensores da terra plana podem tornar a sua posição consistente com os indícios da geografia, da astronomia, e das fotografias de satélites, construindo relatos forçados sobre conspirações a fim de encobrir a verdade, sobre os efeitos do espaço vazio sobre as câmeras, e assim por diante. Mas isso não mostra que precisamos levar a sério o seu “terraplanismo”. Similarmente, muito embora a teoria gravitacional newtoniana possa em princípio ser tornada consistente com todos os indícios contrários adicionando-se várias forças subjacentes e manobras *ad hoc*, isso não é razão para não se acreditar na teoria da relatividade geral.

Certamente que os cientistas praticantes não consideram a SDTI com um bloqueio ao seu acesso à verdade teórica. Eles reconhecem que em princípio podemos sempre inventar explicações alternativas para qualquer corpo de dados; mas eles simplesmente as desconsideram, não levando a sério essas alternativas que precisam invocar planetas escondidos, ou forças subjacentes, ou outras conspirações para encobrir a verdade. Com efeito, os cientistas são ensinados, no curso de seu treino científico, que apenas certos tipos de teoria são candidatos possíveis à verdade; e uma vez que têm dados que excluem todas menos uma *dessas* teorias, eles ignoram bastante satisfeitos todas as outras teorias conspiratórias que permanecem consistentes com os dados. (Talvez o melhor modo de descrever esse aspecto da prática científica seja dizer que os cientistas ignoram todas as teorias que não são suficientemente “simples”; mas se pensarmos assim, não deveríamos pensar na “simplicidade” como uma idéia inata ou intuitiva; ao invés, o tipo relevante de simplicidade é parte daquilo que os cientistas

aprendem quando são treinados como meteorologistas, embriologistas, físicos, e assim por diante.)

Contundo, ainda que os cientistas não considerem a SDTI como um obstáculo sério, muitos filósofos, como eu disse, passam rapidamente da premissa de que diferentes teorias são consistentes com os indícios observacionais à conclusão de que nenhuma delas pode ser considerada como a verdadeira. No entanto, penso que eles apenas dão esse passo porque aceitam que as únicas inferências boas dos dados às teorias são as inferências dedutivamente válidas: eles notam que os dados não podem implicar T dedutivamente se deixarem aberta a possibilidade de que uma teoria inconsistente T' seja verdadeira; e concluem que isso mostra que nunca estamos autorizados a acreditar em T.

No entanto, como vimos em nossa discussão anterior sobre a indução na Seção 1, há boas razões para conceder que outras inferências além das inferências dedutivamente válidas possam ser racionais. Em particular, naquela discussão sugeri que a exigência básica importante poderia ser simplesmente a de que as inferências devessem ser fiáveis e não dedutivamente válidas.

Na verdade, a questão que estamos a tratar agora está intimamente relacionada à nossa discussão anterior acerca da indução. Na seção 1 foquei-me na indução *enumerativa*, na qual partimos de instâncias de um padrão à teoria de que esse padrão vale geralmente. As escolhas teóricas que estamos agora a considerar podem ser pensadas como induções *eliminativas*, nas quais presumimos que a verdade se encontra em uma dentre uma quantidade limitada de teorias (as teorias razoavelmente “simples”), e então usamos as nossas observações para eliminar todas essas teorias menos uma.

A diferença essencial entre essas duas formas de indução é que as induções eliminativas consideram apenas uma quantidade limitada de teorias como candidatas à verdade. Isso poderia fazer como que a indução enumerativa parecesse uma forma mais geral de inferência, uma vez que não se assenta em tal pressuposição. Mas a nossa discussão do “novo problema da indução” de Goodman na Seção 1 mostra de fato que mesmo as induções enumerativas assentam-se numa pressuposição similar: uma vez

que há tantos modos possíveis de projetar padrões observados no futuro, as induções enumerativas são forçadas a restringir as generalizações que consideram como candidatas à verdade a uma quantidade limitada que envolve predicados projetáveis. Por exemplo, as proposições do tipo “Todas as esmeraldas são verdes (amarelas/ vermelhas/ etc.)” são razoavelmente “simples”, e por isso candidatas à verdade, mas as proposições do tipo “Todas as esmeraldas são verdes (azules/ etc.) não são. Alguém que esteja investigando esmeraldas pode então chegar à conclusão natural por notar quais das candidatas à verdade é consistente com as observações feitas até agora.

Dado isso, podemos também considerar todas as induções como essencialmente eliminativas ao invés de enumerativas. Contudo, o problema da fiabilidade surge do mesmo modo na indução eliminativa como na indução enumerativa. O fato de as induções eliminativas não serem logicamente válidas não significa que não sejam fiáveis. Mas permanece a questão de saber se são fiáveis.

Na Seção 1 sugeri que seria aceitável responder essas questões para as induções enumerativas fornecendo indícios (enumerativamente) metaindutivos a favor de sua fiabilidade. Talvez possamos tentar a mesma manobra novamente. Isto é, talvez possamos tomar como indício aquelas ocasiões em que os cientistas escolheram uma teoria “simples” que é consistente com os dados, e então argumentar metaindutivamente que a explicação “simples” do sucesso dessas inferências é que tais induções eliminativas são em geral guias fiáveis para a verdade. Essa manobra obviamente envolve algum elemento de circularidade, mas, como notei na Seção 1, não é claro que esse tipo de circularidade seja viciosa.

Poder-se-ia dizer que esse é o único modo possível pelo qual poderíamos tentar defender a racionalidade da indução eliminativa. O principal ponto que pretendo realçar nesta seção é que a racionalidade da indução eliminativa não requer que ela seja dedutivamente válida. Assim, a SDTI não mostra que tais induções nunca são aceitáveis, e por isso não torna duvidosa a posição realista de que as teorias sobre inobserváveis bem confirmadas possam ser consideradas como descrições verdadeiras da natureza. Como ultrapassar isso e mostrar positivamente que as induções eliminativas *são* racionais talvez seja uma questão demasiado difícil para se resolver aqui.

3.6. A Metaindução Pessimista a partir da Falsidade Passada

Deixe-me retornar ao outro argumento contra o realismo mencionado anteriormente. Esse argumento toma como premissa que as teorias científicas passadas geralmente se mostraram falsas, e então passa indutivamente à conclusão pessimista de que as nossas teorias atuais são, sem dúvida, falsas também.

Há uma profusão de exemplos familiares para sustentar esse argumento. A teoria de Newton do espaço e do tempo, a teoria do flogisto da combustão, e a teoria de que os átomos são indivisíveis já foram teorias científicas amplamente aceitas embora já tenham sido reconhecidas como falsas. Não parece plausível, então, concluir a indução pessimista, que todas as nossas teorias científicas atuais sejam falsas, e que devêssemos, portanto, tomar uma atitude instrumentalista ao invés de realista perante elas?

Esse é um importante e poderoso argumento, mas seria apressado concluir que torne o realismo completamente duvidoso. É importante que a tendência à falsidade seja muito mais comum em algumas áreas das ciências do que em outras. Assim, é relativamente normal que teorias sejam derrubadas na cosmologia, digamos, ou na física de partículas, ou no estudo da evolução dos primatas. Em contraste, as teorias da composição dos diferentes compostos químicos (como, por exemplo, de que a água é feita de hidrogênio e oxigênio), ou as causas de doenças infecciosas (que erupções se devem ao vírus da herpes), ou a natureza dos fenômenos físicos cotidianos (que o calor é o movimento molecular) são caracteristicamente conservadas uma vez aceitas.

Nem precisamos considerar essa frequência de sucesso diferencial de diferentes tipos de teorias como algum tipo de acidente. Ao invés, é o resultado dos indícios necessários estarem mais facilmente disponíveis em algumas áreas da ciência do que em outras. Os paleoantropólogos querem saber quantas espécies de homínídeos estavam presentes na terra há três milhões de anos. Mas os indícios que eles possuem consistem de poucas peças de dentes e ossos. Por isso não é surpreendente que as descobertas de novos sítios de fósseis os faça frequentemente mudar suas perspectivas. O mesmo se aplica em ampla escala à cosmologia e à física de partículas. Os cientistas nessas áreas querem responder questões mais gerais acerca do muito pequeno e do muito distante. Mas os indícios que eles possuem derivam de um domínio limitado de instrumentos

tecnológicos projetados para adentrar nesses reinos. Por isso, mais uma vez, pouco surpreende que as suas teorias permaneçam no nível de hipóteses. Em contraste, nas áreas onde os indícios adequados estão disponíveis, como a química e a medicina, não há barreira correspondente à passagem que a ciência faz das hipóteses às conclusões seguras.

A moral da estória é que o realismo é mais defensável em algumas áreas da ciência do que em outras. Em algumas questões científicas indícios seguros estão disponíveis e nos permite averiguar certas teorias, como a teoria de que a água é composta de moléculas de H_2O , como sendo a verdade literal acerca da realidade. Em outras áreas os indícios são fragmentados e inconclusivos e fazemos melhor se considerarmos as teorias bem apoiadas, como a teoria de que quarks e leptons são constituintes últimos da matéria, como instrumentos úteis que acomodam os dados existentes, fazendo previsões interessantes e sugerindo linhas ulteriores de pesquisa.

À primeira vista, poderia parecer que isso é uma vitória do instrumentalismo sobre o realismo. Pois os instrumentalistas não aceitam que deveríamos ser realistas acerca das coisas observáveis, e apenas recomendam o instrumentalismo para as teorias incertas acerca de objetos inobserváveis? Em contraste, a posição que alcançamos não tem peso algum sobre a distinção entre o que é observável e o que não é. Em particular, ela defende que a metaindução pessimista não mostra que a falsidade é o destino natural de todas as teorias sobre inobserváveis, mas apenas que há uma linha na categoria das teorias acerca de inobserváveis entre as teorias que se pode esperar que se mostrem falsas e aquelas cujas reivindicações à verdade sejam seguras. Assim, nossa posição atual não é um instrumentalismo dogmático sobre todos os inobserváveis, mas apenas a perspectiva indisputável de que deveríamos ser instrumentalistas sobre a subclasse de teorias que não são apoiadas pelos indícios adequados.

4. CONFIRMAÇÃO E PROBABILIDADE

4.1. A noção de Confirmação

No final da seção anterior argumentei que a história da ciência nos dá razões para sermos cautelosos em nosso comprometimento com certas teorias científicas. Em pelo menos algumas áreas da ciência os indícios a favor mesmo das melhores teorias são geralmente fragmentários e inconclusivos, com a conseqüência de que deveríamos esperar que tais teorias se mostrassem falsas.

Seria bom conseguir dizer mais sobre o grau ao qual um dado corpo de indícios apóia dada teoria. Isto é, seria bom termos uma abordagem quantitativa da relação entre indício e teoria. Os filósofos têm procurado desenvolver tais abordagens sob o nome de “teoria da confirmação”. Eles procuram entender em que medida diferentes corpos de indícios “confirmam” diferentes teorias. Se uma teoria é altamente confirmada pelos indícios disponíveis, então podemos razoavelmente confiar que é verdadeira; mas se tiver um baixo grau de confirmação, então deveríamos moderar a nossa confiança nela.

No entanto, essa noção intuitiva de confirmação é menos simples do que parece. Introduzirei algumas das dificuldades descrevendo dois conhecidos paradoxos que qualquer teoria da confirmação tem de dar conta.

4.2. O Paradoxo dos Corvos

Presumamos que haja uma relação de confirmação de acordo com a qual E confirma T, em que E é algum corpo de indícios e T alguma teoria. Então, certamente parece natural fazer as seguintes duas suposições sobre a confirmação:

(1) Se $E = (Fa \ \& \ Ga)$ e $T = \text{Todos os } Fs \text{ são } Gs$, então *E confirma T*.

(Essa primeira suposição simplesmente diz que as generalizações são confirmadas por suas instâncias.)

(2) Se E confirma T, e T é logicamente equivalente a S, então *E confirmam S*.

Como já disse, essas duas suposições parecem ser altamente indisputáveis. Mas pode-se mostrar facilmente que geram um quebra-cabeça.

Note primeiro que as seguintes duas generalizações são logicamente equivalentes:

(L) Todos os corvos são pretos.

(M) Todas as coisas não-pretas são não-corvos.

Agora tome como indício uma observação de que:

(I) Aquela coisa branca é um sapato.

Uma vez que (I) é uma instância de uma coisa não-preta que é um não-corvo, então a suposição (1) nos diz que (I) confirma (M).

Mas se agora juntarmos isso com o fato de que (M) é logicamente equivalente a (L), então a suposição (2) nos diz que (I) confirma (L).

No entanto isso parece absurdo. Pois (L) é a afirmação de que todos os corvos são pretos, e certamente não podemos confirmá-la apenas por observar que alguma coisa branca é um sapato.

Algo parece estar errado algures. Mas é difícil ver onde. Pois dificilmente pode haver algo de errado com a suposição (2) – proposições logicamente equivalentes fazem exatamente as mesmas afirmações sobre o mundo, de modo que é difícil ver como uma porção de dados poderia apoiar uma proposição sem que com isso apóie a outra. E a suposição (1) parece quase que óbvia – se algo é sempre confirmado por algo, as generalizações são certamente confirmadas por suas instâncias.

(Alguns poderiam pensar que a falha no raciocínio está na suposição (1). Pois não é uma lição do novo problema da indução de Goodman precisamente que $Fa \ \& \ Ga$ não podem sempre confirmar $\forall x (Fx \rightarrow Gx)$? Goodman mostra que, a menos que restrinjamos F e G a predicados “projetáveis”, há muitos Fs e Gs para todas as generalizações serem confirmáveis por suas instâncias. No entanto, não penso que isso ajude com o paradoxo dos corvos, dado que não há qualquer coisa particularmente bizarra (como no caso do verdul) com os predicados usados para formulá-lo, designadamente, “preto”, “corvo”, “não-preto” e “não-corvo”. É certamente verdadeiro que o argumento de Goodman mostra que (1) não é aceitável se formulado sem a qualificação acima. Mas o paradoxo será gerado ainda que (1) se restrinja apenas a aplicação a predicados “projetáveis”.)

4.3. O Paradoxo da Inclusão

Vamos ao segundo paradoxo. Eis duas outras suposições que parecem perfeitamente óbvias.

(3) Se T acarreta E, então E *confirma* T.

(Essa é apenas a idéia de que uma teoria é confirmada se as suas conseqüências são observadas com sendo verdadeiras.)

(4) Se E confirma T, e T acarreta P, então E *confirma* P.

(Essa é apenas a idéia de que, se algum indício lhe permite acreditar em alguma teoria, então ele lhe permite acreditar naquilo que se segue dela.)

Mas agora tome uma teoria N – a teoria gravitacional newtoniana, digamos – e *qualquer* conseqüência M que ela acarreta – os planetas se movem em elipses. Considere então *qualquer* outra proposição Q que queiras – a lua é feita de queijo fresco. Uma vez que N (a teoria newtoniana) acarreta M (as órbitas elípticas), por hipótese, N & Q (a teoria newtoniana *mais* a lua ser feita de queijo fresco) também acarreta M. Portanto, por (3)

(a) M confirma N & Q.

Mas

(b) N & Q acarreta Q trivialmente,

de modo que por (4), aplicado a (a) e (b), segue-se que M (as órbitas elípticas) *confirma* Q (a lua é feita de queijo fresco). Mas isso significa que algo que se segue de uma teoria – os planetas se moverem em elipses – confirma qualquer outra teoria que você queira – a lua é feita de queijo fresco. E isso certamente é absurdo.

Essa conclusão absurda, todavia, segue-se das suposições aparentemente incontestáveis (3) e (4). Uma vez mais, é difícil ver onde se encontra o erro de nosso raciocínio.

Esse paradoxo é chamado de paradoxo da “inclusão” pois “inclui” uma hipótese arbitrária (a lua é feita de queijo fresco, no exemplo acima) na teoria com que começamos (a mecânica newtoniana). Uma reação inicial comum é a de que a suposição (3) está errada. Os movimentos dos planetas realmente confirmam a teoria newtoniana e que a lua é feita de queijo fresco? Mas vou argumentar que isso é de fato algo sensato de se supor, e que a suposição (4) é que é realmente responsável pelo paradoxo da inclusão. No entanto, antes de explicar como penso que o paradoxo da inclusão (e o paradoxo dos corvos) deva ser resolvido, será necessário fazer uma digressão e explicar algumas idéias sobre *probabilidade*.

4.4. As Interpretações da Probabilidade

A noção de probabilidade pode ser entendida de diversas formas diferentes. Em particular, como eu disse na Seção 1, há noções tanto objetivas quanto subjetivas da probabilidade. Mas há uma coisa que une todas as diferentes noções de probabilidade, a saber, elas satisfazerem os *axiomas do cálculo de probabilidades*.

Esses axiomas são normalmente expressos com se segue:

- (1) $0 \leq \text{Prob}(p) \leq 1$, para qualquer proposição p .
- (2) $\text{Prob}(p) = 1$, se p é uma verdade necessária.
- (3) $\text{Prob}(p) = 0$, se p é impossível.
- (4) $\text{Prob}(p \text{ ou } q) = \text{Prob}(p) + \text{Prob}(q)$, se p e q são mutuamente exclusivas.

Qualquer maneira de atribuir números às proposições de modo a satisfazer esses axiomas constitui uma interpretação do cálculo de probabilidades. Vamos nos concentrar em particular no contraste entre as interpretações subjetiva e objetiva da probabilidade.

A interpretação subjetiva considera a probabilidade de p como sendo uma medida da força com a qual acreditamos em p . Mais especificamente, para qualquer pessoa X , a interpretação subjetiva iguala a probabilidade de p para X com o grau ao qual X acredita em p .

Alguns subjetivistas extremos argumentam que essa é a única noção de probabilidade que precisamos. Mas a maioria dos filósofos que reconhece as probabilidades subjetivas também reconhece as probabilidades objetivas. As probabilidades objetivas se aplicam especificamente a proposições que afirmam que certo tipo de resultado ocorrerá sob certo tipo de teste repetível, como por exemplo, de que certo tipo de moeda virará cara quando arremessada. E nesse tipo de contexto um enunciado de probabilidade objetiva especifica que a quantidade de testes desse tipo *tende* a produzir o resultado em questão. Esse tipo de tendência é exibido pela frequência com que o resultado ocorre – por exemplo, com que frequência moedas como essa viram cara.

Deveria estar claro que essas interpretações, a subjetiva e a objetiva, nos dão noções diferentes de probabilidade. Um grau de crença subjetiva é uma coisa, e uma tendência objetiva é outra. Não há garantia que quaisquer expectativas subjetivas das pessoas devam corresponder às tendências objetivas; mas haveria probabilidades objetivas de átomos decaindo, ainda que nunca houvesse quaisquer seres humanos para formar graus de crença. Examinemos mais detalhadamente essas duas noções.

4.5. Probabilidades Subjetivas

A suposição central da interpretação subjetiva é a de que a crença ocorre em *graus*. Normalmente pensamos na crença como algo que se possa ter ou não. Mas considere a atitude de alguém que leva uma sombrinha e um protetor solar para um passeio. Essa pessoa acredita que vai chover ou não? A resposta natural é que essa pessoa tem alguma expectativa de que essa proposição seja verdadeira, e alguma de que não seja. Ou, considere a atitude de um diretor de uma companhia que dá dinheiro tanto para o Partido Trabalhista quanto para o Partido Conservador antes das eleições. Novamente, parece natural dizer que o diretor da companhia tem um grau de crença positivo de que cada um dos partidos vencerá. (Alguns objetam à idéia de “graus de crença” porque pensam nas crenças como atitudes definidas, a favor ou contra, frente a proposições. Caso se prefira, poder-se-ia pensar em termos de graus de expectativa ao invés de crença. Isso não fará diferença para a discussão que se segue.)

Uma coisa é defender que as crenças ocorrem em graus. Outra é mostrar que podemos atribuir valores definidos entre 0 e 1 a esses graus. Mas a teoria subjetivista precisa mostrar isso, uma vez que os graus de crença terão que se igualar a tais valores caso tenham alguma chance de satisfazer os axiomas da probabilidade.

No entanto, isso não é necessariamente tão forçado quanto parece à primeira vista. A maneira óbvia de atribuir um valor ao grau de crença de alguém é ver que mínima vantagem a induziria a apostar em p . Se você está disposto a apostar apenas R\$ N uma vez que seu oponente oferece R\$ M ou mais, levando tudo quem ganhar caso p se mostre verdadeira, então isso defensavelmente mostra que o seu grau de crença em p é $N/(N + M)$.

É verdade que algumas pessoas detestam apostar *per se*. E em tais casos, as vantagens que as induzirão a apostar sobreestimarão os seus graus de crença. Por exemplo, você pode estar convencido de que esteja apostando numa moeda não viciada, e por isso atribui uma probabilidade de 50-50 de sair cara, mas pode estar desinclinado a arriscar seus valiosos R\$ 10 quando aposto R\$ 40 ou mais. O teste sugerido no parágrafo anterior indicaria que o seu grau de crença em sair cara é 0,2 e não 0,5. Mas talvez nesse tipo de caso um investigador pudesse ainda descobrir o seu real grau de crença pedindo-lhe que escolha vantagens para apostar em p , sem lhe dizer de que modo você vai apostar, ou o quão boa será. Nessa situação, qualquer aversão à aposta anularia e deixaria a vantagem escolhida expressando o seu real grau de crença.

A despeito dessas sugestões engenhosas, você pode ter a impressão de que é fantasioso supor que haja graus numéricos precisos de crença para todas as proposições. Certamente que não faz diferença se o meu grau de crença de que X vencerá as próximas eleições é de 0,3456 ao invés de 0,3457. Mas o defensor da interpretação subjetivista pode razoavelmente responder que a postulação de graus exatos de crença é uma *idealização* útil que facilita a nossa teorização e que em nada nos prejudica entendemo-la assim. Para fins de comparação, considere o modo em que os físicos supõem que os objetos físicos, como pedras e planetas, têm massas e tamanhos precisos. Isso nunca é estritamente verdadeiro, uma vez que tais objetos sempre perdem e ganham moléculas. Mas a ficção das quantidades precisas é extremamente útil na física e não ilude ninguém.

A maneira normal de se estabelecer que os graus de crença se conformam com os axiomas da probabilidade é através do “argumento da aposta holandesa”, que mostra que alguém cujos graus de crença violam os axiomas (1)-(4) podem ser induzidos a fazer apostas manifestamente irracionais.

Suponha que você atribua uma probabilidade de 0,8 a “Choverá hoje”, e um grau de crença de 0,7 a “Não choverá hoje”. Os seus graus de crença violam assim os axiomas da probabilidade. (Isso porque a proposição de que choverá ou não é uma verdade necessária, e por isso, para satisfazer ao axioma (2) precisa ter grau de crença 1; mas também o é a disjunção das proposições exclusivas de que choverá hoje e de que não choverá hoje, e por isso, para satisfazer ao axioma (4) precisa ter um grau de crença igual à soma dos graus de crença dessas proposições separadas; essa soma, no entanto, é 1,5 e não 1.)

Note também que, porque você tem esses graus de crença, estará disposto a apostar os seus R\$ 8 contra os meus R\$ 2 na proposição de que choverá hoje; e você estará disposto a apostar os seus R\$ 7 contra meus R\$ 3 na proposição de que não choverá hoje. Mas esse é um par de apostas bastante tolo, uma vez que você tem a garantia de que vai perder R\$ 5 aconteça o que acontecer.

É provável, de acordo com o exemplo anterior, que as pessoas estejam vulneráveis às “apostas holandesas” se e só se os seus graus de crença não se conformem aos axiomas da probabilidade. Uma vez que parece claramente irracional ter graus de crença que possam lhe conduzir a fazer coisas que estão condenadas ao fracasso, isso mostra que o grau de crença de todo mundo deve racionalmente se conformar aos axiomas da probabilidade.

Note que a conclusão desse argumento é apenas que os graus de crença de uma pessoa racional devem se conformar aos axiomas da probabilidade, e não que os graus de crença de todos de fato se conformarão. Afinal, a maioria das pessoas provavelmente tem graus de crença que não somam 1 para pelo menos alguns conjuntos exclusivos e exaustivos de proposições. Portanto, o máximo que a interpretação subjetivista pode dizer é que graus *racionais* de crença são uma interpretação do cálculo de probabilidades, e não que os graus efetivos de crença o são.

Note também que ao passo que o argumento da aposta holandesa mostra que os nossos graus de crença devam se conformar ao cálculo de probabilidades, daí não se segue que você deva atribuir um valor *particular* à proposição de que choverá hoje. Você pode atribuir 0,7, ou 0,1 ou, 0,435, ou qualquer valor que você queira a essa proposição, desde que esse grau de crença atribuído a “Não choverá hoje” seja 1 menos esse valor. O argumento da aposta holandesa apenas mostra que os seus graus de crença têm de ser “coerentes” (isto é, têm de algum modo de satisfazer os axiomas (1)-(4)); além disso, é uma questão de escolha subjetiva que graus de crença você tem. Diferentes pessoas podem atribuir diferentes “probabilidades subjetivas” à mesma proposição. A exigência é apenas que para cada pessoa os valores em questão satisfaçam os axiomas da probabilidade; mas esses valores podem ser completamente diferentes para diferentes pessoas.

É nesse último ponto que a maioria das pessoas pensa que precisamos de outra noção de probabilidade – a probabilidade objetiva – para cobrir a idéia de que tais moedas (ou dados, ou átomos de radium) tenham certas tendências para virar cara (virar seis, decair). Pois essas tendências objetivas presumivelmente têm valores objetivos definidos, ainda que pessoas diferentes tenham graus de crença diferentes no resultado relevante.

4.6. Probabilidades Objetivas

Há duas maneiras rivais de se pensar sobre a probabilidade objetiva, a teoria *frequencista* e a teoria *propensista*. Considerá-las-ei nessa ordem.

4.6.1. A Teoria Frequencista

A maneira tradicional de dar sentido às probabilidades objetivas é igualá-las às freqüências relativas dos resultados. Assim, igualamos a probabilidade p do resultado R (cara, seis, decaimento) na situação S (arremesso de moeda, lançamento de dado, átomo de radium) a:

a quantidade de R s/a quantidade total de S s.

Note que isso apenas nos permite atribuir probabilidade a resultados que acontecem em situações repetíveis em que temos uma quantidade de Ss, e não a todas as proposições, como na teoria subjetivista. Mas isso não é uma crítica, uma vez que é defensável que a noção de probabilidade objetiva se aplica apenas a tais situações repetíveis, e não a proposições de rara ocorrência como a de que o Príncipe Edward se casará neste ano.

Um problema óbvio que a definição acima enfrenta é o de saber qual o “número total” de testes S deveríamos considerar. Normalmente não pode ser os resultados efetivos do tipo S, uma vez que normalmente serão finitos em quantidade. O problema aqui é que sabemos (uma vez que se segue dos axiomas da probabilidade) que há sempre uma probabilidade diferente de zero de que a frequência relativa após N testes será diferente de p , já que N é finito. Por exemplo, é inteiramente possível (e de fato altamente provável) que 1.000 arremessos de moedas com uma probabilidade objetiva 0,5 de virar cara seja algo além das exatas 500 caras. Portanto, não há garantia afinal de que a frequência relativa numa quantidade finita qualquer de testes será igual à probabilidade objetiva.

Por causa disso, a teoria frequentista define comumente as probabilidades não em termos de frequências em conjuntos infinitos de testes, mas em termos da proporção dos Rs que ocorreriam *caso* o teste S fosse repetido infinitamente.

Esse apelo às sequências infinitas de resultados levanta uma dificuldade técnica. Pois a noção de uma *proporção* de Rs numa sequência infinita de Ss não faz sentido. Se arremessarmos uma moeda uma quantidade infinita de vezes, então haverá uma quantidade infinita de caras e uma quantidade infinita de coroas. Portanto, a proporção de caras na quantidade total de arremessos é infinito dividido por infinito, o que não faz sentido. O modo de contornar essa dificuldade é igualar a probabilidade com o *limite* da frequência relativa *finita* de Rs no primeiro n Ss, em que n se torna cada vez maior. Mais precisamente, podemos dizer que a frequência relativa de m Rs no primeiro n Ss *tende* ao limite p (e então igualar a probabilidade objetiva com esse p) se

para qualquer ϵ , ainda que pequeno, haja um N, tal que, para todo $n > N$, $- \epsilon < m/n - p < + \epsilon$.

(Essa é apenas a idéia matemática padrão de limite – um número tal que, para qualquer região minúscula em sua volta, a frequência relativa eventualmente se manterá nessa região uma vez que você tenha percorrido a sequência o bastante.)

No entanto, ainda que a teoria frequencista possa lidar com esse problema técnico suscitado pelas sequências infinitas, muitos filósofos ainda se sentem desconfortáveis em definir as probabilidades em termos de fatos hipotéticos acerca daquilo que *aconteceria se S* ocorresse infinitamente. Uma vez que a maioria dos *Ss*, como arremessos de moedas, lançamentos de dados, ou decaimentos atômicos, efetivamente não ocorrem infinitamente, isso significa que estamos tentando definir as probabilidades objetivas em termos de fatos não existentes e imaginários. Isso convenceu muitos filósofos a procurar uma abordagem alternativa às probabilidades objetivas.

4.6.2. A Teoria Propensista

A teoria propensista da probabilidade objetiva abandona a idéia de frequências relativas em testes repetidos, e defende que deveríamos simplesmente tomar a noção de probabilidade como uma noção *primitiva* que mede a força da *propensão* de cada *S* particular produzir *R*. Os teóricos da propensão usam normalmente o termo “chance” para referir essa quantidade. Assim, quando dizem que a chance desta moeda de virar cara quando eu arremessá-la é de 0,4, querem dizer simplesmente que essa combinação particular da moeda e do arremessador tem uma tendência de 0,4 para produzir caras.

A teoria propensista tem a desvantagem de não definir a probabilidade, mas simplesmente a toma como primitiva. Por outro lado, tem a vantagem de não precisar recorrer às sequências infinitas não-existentes da teoria frequencista. Qual dessas duas teorias você prefere dependerá principalmente de se você pensa que as sequências infinitas são um preço a se pagar por uma definição explícita.

À primeira vista poderia parecer que a teoria propensista terá mais dificuldades em explicar como *descobrimos* as probabilidades objetivas do que a teoria frequencista. Pois certamente o nosso conhecimento das probabilidades objetivas advém da observação das frequências. Contudo, a teoria propensista parece negar qualquer conexão entre as probabilidades e as frequências.

Os teóricos da propensão, no entanto, podem retorquir que reconhecem uma conexão perfeitamente legítima entre as probabilidades objetivas e as frequências, ainda que não seja uma conexão que defina a primeira em termos da última. Pois podem mostrar que é um teorema do cálculo de probabilidades que

numa sequência de n testes, cada um com a probabilidade p do resultado R , a *probabilidade* de que a frequência relativa de R s será próxima a p pode ser tornada tão alta quanto você queira por tornar n grande o bastante.

Isso não produz uma definição de probabilidade em termos de frequência, uma vez que *usa* a noção de probabilidade ao explicar a conexão entre probabilidade e frequência (note a “probabilidade” enfatizada no enunciado do teorema). Mas é ainda uma conexão que nos permite tomar as frequências como indício das probabilidades.

Na verdade, eles não fornecem indícios seguros, uma vez que mesmo para um n amplo é apenas *provável* que a frequência será próxima à probabilidade, e não certa. Mas esse problema (“o problema da inferência estatística”) não é um problema apenas para a teoria propensista. Afinal, mesmo os teóricos frequentistas têm que descobrir as probabilidades com base em frequências *finitas* (uma vez que nunca observam frequências infinitas). Portanto, eles também enfrentam de que é no máximo provável, e não certo, que a probabilidade objetiva (isto é, para as teorias frequentistas, a frequência no limite finito) será próxima da frequência observada.

Esse problema da inferência estatística é apenas um aspecto da filosofia da probabilidade que não podemos tratar aqui. O nosso tratamento tanto da probabilidade objetiva quanto da subjetiva apenas tocou a superfície desses tópicos. Mas já temos o bastante para continuar a nossa discussão da teoria da confirmação.

Deixe-me fazer outro comentário antes de retornar a principal linha argumentativa. Até agora eu não disse qualquer coisa acerca da conexão entre as probabilidades objetiva e subjetiva. Não há dúvidas de que essas são noções distintas, como mostrei anteriormente. Mas isso não significa sejam desconexas. Mais especificamente, o seguinte princípio resume uma importante conexão:

Se você sabe que a probabilidade objetiva de R no tempo t é p , então em t o seu grau de crença em R dever ser igual a p .

Essa idéia parece demasiado óbvia para valer a pena mencioná-la. Naturalmente, se sei que esta moeda agora uma probabilidade objetiva de 0,5 de virar cara, tornarei o meu grau de expectativa para esse resultado igual a 50 por cento. No entanto, vale a pena observar, antes de deixarmos esse tópico, que nenhuma das teorias da probabilidade objetiva e subjetiva delineada acima oferece qualquer explicação de por que esse princípio é verdadeiro. Uma vez mais, há mais coisas que dizem respeito à probabilidade do que somos capazes de tratar aqui.

4.7. A Teoria Bayesiana da Confirmação

Volto-me agora ao tópico da teoria da confirmação. No restante desta seção irei me concentrar na teoria da confirmação bayesiana. Os bayesianos são filósofos que pensam que podemos usar a noção de probabilidade subjetiva para explicar a relação de confirmação. Essa não é necessariamente a única maneira de se pensar sobre a confirmação. Mas o bayesianismo oferece uma maneira poderosa e uniforme de se pensar acerca dos problemas da confirmação. Em particular, como veremos, fornece soluções naturais aos dois paradoxos da confirmação descritos anteriormente.

A suposição inicial feita pela teoria bayesiana da confirmação é que a nossa atitude perante as teorias são medidas pelas probabilidades subjetivas que atribuímos a elas. Assim, se acredito completamente numa teoria, atribuo-lhe uma probabilidade subjetiva 1; ao passo que se a considero uma especulação arriscada, atribuo-lhe uma probabilidade subjetiva próxima a 0.

Os bayesianos então dizem que um corpo de indícios E *confirma* uma teoria T se ter conhecimento de E fizer com que as pessoas *umentem* a probabilidade que atribuem a T . (No restante da seção tonderei a omitir a qualificação “subjetiva”; a menos que eu diga o contrário, “probabilidade” significará “probabilidade subjetiva”.)

A fim de desenvolver mais a teoria bayesiana, precisamos da noção de probabilidade condicional. A *probabilidade condicional de A dado B* (escrito “Prob (A/B)”) é definida como o quociente da Prob (A e B)/Prob (B), e pode ser pensada como

a probabilidade de A dada a suposição de que B é verdadeira. Para ver por quê, note que $\text{Prob}(B)$ é uma medida da probabilidade de B acontecer, ao passo que $\text{Prob}(A \text{ e } B)$ é uma medida da probabilidade de A *também* acontecer quando B acontece. Assim, se dividirmos $\text{Prob}(A \text{ e } B)$ por $\text{Prob}(B)$ obteremos a medida da probabilidade de A acontecer *dado* que B aconteceu.

Considere agora o caso em que E é um indício possível e T é uma teoria. $\text{Prob}(T/E)$ é, então, a probabilidade de T dada a suposição de que E é verdadeiro. Os bayesianos, portanto, argumentam que quando você tem conhecimento de E, aumentará para esse número a probabilidade que atribui a T. Assim, para os bayesianos E *confirmará* T no sentido de que a descoberta de E aumentará a probabilidade que atribuímos a T, se e só se a $\text{Prob}(T/E)$ for maior que $\text{Prob}(T)$. (Na verdade essa alegação é menos simples do que parece à primeira vista. Mas doravante a presumirei. Para mais discussões veja a leitura adicional.)

Podemos dizer mais sobre *quando* E confirmará T se prestarmos atenção no *Teorema de Bayes*, originalmente descoberto pelo clérigo inglês Thomas Bayes no século XVIII. Esse teorema se segue diretamente da definição de probabilidade condicional. De acordo com essa definição a $\text{Prob}(T/E) = \text{Prob}(T \text{ e } E)/\text{Prob}(E)$, em que $\text{Prob}(E/T) = \text{Prob}(T \text{ e } E)/\text{Prob}(T)$. Juntando as duas podemos derivar

$$\text{Prob}(T/E) = \text{Prob}(T) \times \text{Prob}(E/T)/\text{Prob}(E).$$

Esse é o teorema de Bayes. A sua significância é que ele nos diz que a $\text{Prob}(T/E)$ é maior do que a $\text{Prob}(T)$ – isto é, E confirma T – se e só se a $\text{Prob}(E/T)$ for maior do que a $\text{Prob}(E)$. Isso é o que esperaríamos pré-teoricamente. Pois diz que E confirma T na medida em que E é provável dado T, e improvável do contrário. Em outras palavras, se E em si é bastante surpreendente (como o a curvatura da luz nas adjacências do sol) embora ao mesmo tempo seja aquilo que você esperava dada a sua teoria T (a teoria geral da relatividade), então E faria com que você aumentasse consideravelmente o seu grau de crença em T. Por outro lado, se E não é mais provável dado T do que seria dada outra teoria qualquer, então observar E não fornece apoio extra a T. O movimento das marés, por exemplo, não é uma razão forte a favor da teoria geral da relatividade, muito embora

seja prevista por ela, uma vez que é também prevista pela teoria newtoniana da gravitação.

4.8. Os Paradoxos Resolvidos

Consideremos agora como essa abordagem bayesiana da confirmação lida com os paradoxos da confirmação.

4.8.1. O Paradoxo dos Corvos

Primeiro o paradoxo dos corvos. As suposições que geram o paradoxo, lembre-se, são (1) que as generalizações são confirmadas por suas instâncias, e (2) que a confirmação vale igualmente para proposições logicamente equivalentes. A resposta bayesiana padrão a esse paradoxo é aceitar ambas as suposições, e com isso a conclusão aparentemente absurda de que um sapato branco confirma que todos os corvos são pretos. Mas os bayesianos então explicam essa aparência de absurdidade dizendo que um sapato branco confirma *pouquíssimo* essa hipótese em comparação com a confirmação obtida de um corvo preto.

Deixe-me usar imagens simples para ilustrar o ponto. Suponha que você inicialmente pense que cerca de $1/5$ dos objetos físicos sejam pretos, e que cerca de $1/10$ sejam corvos. (Isso não é muito realista, mas mantenhamos as imagens simples.) Então, na falta de quaisquer panoramas especiais acerca das cores dos corvos, a probabilidade que você atribui ao próximo objeto que você ver como sendo um corvo preto será $1/50$, e como sendo um não-corvo preto será $36/50$ (e similarmente como sendo um corvo não preto $4/50$ e um não-corvo não-preto $9/50$).

Considere agora a probabilidade condicional de um corvo preto e um não-corvo não-preto de acordo com a suposição (T) de que todos os corvos são pretos. Essa suposição tenderá a aumentar a probabilidade que você atribui a ambas observações simplesmente porque ela diminui a probabilidade de que você verá um corvo não-preto de $4/50$ para zero. Suponha que essa probabilidade condicional para um corvo não-preto seja $2/50$, para um não-corvo não-preto $38/50$ (e para um não-corvo preto $10/50$).

Podemos agora aplicar o teorema de Bayes. A probabilidade inicial de um corvo preto é $1/50$, ao passo que a probabilidade condicional dado T é $2/50$. Portanto, qualquer que seja a probabilidade inicial que você atribua à hipótese de que todos os corvos são pretos (equivalentemente, para todos os não-pretos são não-corvos), o teorema de Bayes nos diz que uma observação de um corvo preto a dobrará. Em contraste, quando a probabilidade inicial de um não-corvo não-preto é $36/50$, a probabilidade condicional de acordo com T é apenas $38/50$. Assim, a observação de um sapato branco aumentará nosso grau de crença na hipótese em $2/36$. O ponto é que a hipótese de que todos os corvos são pretos torna a observação de um corvo preto significativamente menos surpreendente do que se fosse o contrário. Ao passo que a observação de um não-corvo não-preto, não tão surpreendente, torna-se apenas marginalmente menos surpreendente de acordo com a hipótese de que todos os corvos são pretos. Assim, corvos pretos confirmam muito a hipótese; e sapatos brancos a confirmam apenas de maneira escassa.

Talvez seja surpreendente descobrir que sapatos brancos dêem *algum* apoio à hipótese de que todos os corvos são pretos, ainda que apenas um pouco. Mas podemos ver que, com imagens realistas, esse apoio seria tão minúsculo que seria bastante estranho dizer, num contexto comum, que um sapato branco nós dá alguma razão para acreditar que todos os corvos são pretos. É assim que os bayesianos lidam com o paradoxo dos corvos: não negam que sapatos brancos confirmem que todos os corvos são pretos; apenas mostra que a confirmação é tão pouco que não faria diferença num contexto comum.

4.8.2. O Paradoxo da Inclusão

Passemos ao paradoxo da “inclusão”. Lembre-se que as suposições aqui são (1) que as teorias são confirmadas pela observação de algo que acarretam, e (2) que qualquer indício que confirme uma teoria também confirma as suas conseqüências. Muitas pessoas, como mencionei anteriormente, pensam que tem de haver algo errado com (1), uma vez que ela permite que uma teoria mais uma parte “inclusa” (a teoria newtoniana mais a lua é de queijo fresco, digamos) sejam confirmadas pelas previsões da teoria original (os planetas se movem em elipses), o que parece estranho.

Os bayesianos, no entanto, estão comprometidos com (1). Pois se alguma T acarreta E, então a $\text{Prob}(E/T) = 1$. Assim, na medida em que E não é em si necessariamente verdadeira, com uma probabilidade incondicional 1, E tem de confirmar T pelo teorema de Bayes.

Mas os bayesianos mostram que isso é consistente com E confirmando apenas T no sentido de que aumenta a probabilidade de alguma parte de T, deixando a probabilidade do restante de T intocada. Assim, por exemplo, os bayesianos diriam que o movimento dos planetas confirma apenas a teoria newtoniana mais a lua é de queijo fresco no sentido de que aumenta a probabilidade da própria teoria newtoniana, sendo irrelevante para a parte do queijo fresco da hipótese conjuntada.

De acordo com isso, os bayesianos negarão (2). Pois, quando um indício confirma uma teoria apenas no sentido de que aumenta a probabilidade de parte dela, deixando o resto intocado, então esperaríamos que o indício confirme as conseqüências apenas daquela parte, e não da outra. Assim, enquanto que o movimento dos planetas confirma a tese conjuntada da teoria newtoniana mais a lua ser feita de queijo fresco, não confirma a conseqüência de que a lua é feita de queijo fresco, ou o que se siga dela.

Essa via bayesiana não oferece apenas uma solução natural ao paradoxo da inclusão, ela também ajuda a pensar na relação entre teoria e observação de um modo geral. Grande parte da filosofia da ciência recente tem inferido, da observação de Duhem-Quine, que suposições teóricas só geram previsões com a ajuda de hipóteses auxiliares, que a relação entre a teoria e indícios é irremediavelmente *holista*, no sentido de que é sempre a totalidade de nossas crenças sobre o mundo que é confirmada ou refutada por indícios. Mas a abordagem bayesiana mostra que ainda que as previsões sejam geradas por uma conjunção de suposições, esses indícios podem apoiar diferentes elementos da conjunção em diferentes graus.

4.9. Problemas para o Bayesianismo

Não deveríamos nos esquecer de que a teoria bayesiana da confirmação é derivada da noção de probabilidade *subjetiva*. Como notei anteriormente, não há algo na

idéia de probabilidade subjetiva que assegure que diferentes pessoas atribuirão as mesmas probabilidades subjetivas e condicionais a algum conjunto de proposições, não obstante cada uma delas possa organizar “internamente” suas próprias probabilidades de modo tal que satisfaça o cálculo de probabilidades. Pode tal noção de probabilidade subjetiva de fato fornecer uma base satisfatória para a noção aparentemente objetiva de quanto as teorias são confirmadas pelos indícios existentes? Certamente que não queremos permitir que eu possa estar correto ao sustentar que um indício mostra relativamente que a teoria tem probabilidade de 0,8, enquanto que você possa estar igualmente correto ao pensar que ela tem probabilidade de 0,2.

Diferentes bayesianos dão respostas diferentes a esse desafio. Alguns simplesmente dizem que o bayesianismo é apenas uma teoria de como mudar as probabilidades que você atribui, e não acerca das probabilidades com as quais você deve começar ou terminar. Dessa perspectiva, o teorema de Bayes nos mostra como atualizar nossas probabilidades subjetivas dado que começamos com certas probabilidades iniciais condicionais e incondicionais; mas isso não diz qualquer coisa sobre quais devem ser essas probabilidades iniciais e, portanto, sobre com que probabilidades finais devemos terminar. Não há qualquer coisa de errado com que eu termine pensando que a teoria tem relativamente probabilidade 0,8 enquanto que você pense que ela tem probabilidade 0,2, não obstante ambos tenhamos alcançado esse ponto final atualizando nossas probabilidades iniciais em resposta os indícios da maneira exigida pelo teorema de Bayes.

Muitos bayesianos, no entanto, acham preocupante a possibilidade de tal divergência, e por isso oferecem uma resposta mais ambiciosa. Eles dizem que, sejam quais forem os nossos graus de crença, o teorema de Bayes garantirá convergência de opinião. A idéia é que, dados indícios suficientes, eventualmente todos terminarão com as mesmas probabilidades, ainda que tenham pontos de partida diferentes. Há diversos teoremas da teoria da probabilidade que mostram que, dentro dos limites, as diferenças nas probabilidades iniciais serão “rejeitadas”, no sentido de que os indícios suficientes e a atualização bayesiana conduzirão efetivamente a graus de crença finais idênticos. Assim, no final, argumentam os bayesianos, não importa se você começa com um grau de crença alto ou baixo na teoria da relatividade – pois, após diversas observações da

curvatura da luz, de desvios gravitacionais para o vermelho, e assim por diante, acabará de alguma maneira acreditando nela num grau próximo ao das outras pessoas.

No entanto, por mais satisfatórios que sejam esses resultados, eles não respondem satisfatoriamente a preocupação filosófica fundamental. Pois não funcionam para *todos* os possíveis graus de crença. Ao invés, presumem que as pessoas em questão, ao passo que são diferentes entre si, obtêm seus graus de crença iniciais de certo domínio. Ao passo que esse domínio inclui a maior parte dos graus de crença iniciais que parecem de todo plausíveis, há, contudo, outros graus de crença iniciais possíveis que são consistentes com os axiomas da probabilidade, mas que não conduzirão à eventual convergência. Assim, por exemplo, os bayesianos não explicam de fato o que está errado com as pessoas que nunca terminarão acreditando na teoria da relatividade porque sempre pensam que é provável que o curso da natureza mude amanhã.

Parece-me que isso mostra que o bayesianismo fornece, na melhor das hipóteses, uma explicação parcial da confirmação. Os bayesianismo nos mostra como nossos graus de crença iniciais restringem o modo em que responderíamos a novos indícios. Mas ele precisa ser suplementado por uma explicação adicional de por que alguns graus de crença iniciais são objetivamente superiores a outros. Talvez um modo de preencher esse hiato seria apelar ao tipo de “simplicidade” mencionada no Seção 3.6. Mas seria ir longe demais tratar dessa questão aqui.

5. EXPLICAÇÃO

5.1. O Modelo de Cobertura por Lei

Até agora temos tratado principalmente do nosso conhecimento de verdades gerais. Nesta seção focar-me-ei na aplicação desse conhecimento à *explicação*. Tanto na

ciência quanto na vida cotidiana o objetivo da investigação é geralmente encontrar uma explicação para alguns fenômenos intrigantes. Mas o que exatamente é uma explicação? E como o conhecimento de verdades gerais contribui para a nossa capacidade de explicar?

Grande parte de discussão contemporânea da explicação começa com o “modelo de cobertura por lei” de Carl Hempel. Deixe-me primeiro ilustrar esse modelo no caso em que o item a ser explicado é algum evento particular, como, por exemplo, o de que a água em sua caixa d’água congelou terça-feira passada, ou que choveu nesta manhã. De acordo como Hempel, a explicação de tal evento se conforma com o seguinte esquema:

Condições Iniciais: I_1, I_2, \dots, I_n

Leis: L

Evento Explicado: E.

Assim, por exemplo, poderíamos explicar o fato E de que choveu nesta manhã citando as condições iniciais I_1 e I_2 de que havia certo nível de umidade e que a pressão atmosférica baixou a certo nível e a lei L de que tal queda de pressão em tal umidade é sempre seguida por uma precipitação de chuva.

A lei em tal explicação “cobre” as condições iniciais e o evento conseqüente no sentido de que mostra que a seqüência de eventos por trás de uma ocorrência particular é simplesmente uma instância de um padrão geral. O fato que foi explicado, E, é às vezes referido como o “*explanandum*”, e os fatos que fazem a explicação, os Is e L, como “*explanans*”. Note que ao passo que representei a lei envolvida no *explanans* como uma proposição única, L, na maioria dos casos precisaremos de uma conjunção de leis mais simples para ver por que E se segue dos Is relevantes. Por exemplo, precisaríamos tanto da segunda lei de Newton quanto da lei da gravitação para explicar por que um meteoro se move do jeito que move.

Note também que, de acordo com esse modelo de explicação, explicar um evento é a mesma coisa que deduzi-lo de condições iniciais e leis. Dadas as condições iniciais e uma lei que diz que em geral tais condições iniciais são seguidas por um E, então a lógica

apenas nos permite inferir que o *explanandum* ocorre. Porque envolvem *dedução* via *lei*, tais explicações são geralmente chamadas de explicações “dedutivo-nomológicas”, ou explicações “D-N” resumidamente. (Há uma variante do modelo de cobertura por lei que permite leis probabilísticas ao invés de determinísticas, cuja exigência de dedutibilidade é relaxada. Assim, “cobertura por lei” é estritamente um termo mais amplo do que “dedutivo-nomológico”. Mas nos detenhamos por enquanto nos casos dedutivos e deixemos as explicações probabilísticas de lado.)

É bom deixar claro que a idéia de uma explicação “dedutiva” não presume que a lei L possa de algum modo ser “deduzida” de princípios primeiros de maneira *a priori*. Tais leis ainda têm de ser estabelecidas por *indução* de observações passadas de resultados. A idéia é simplesmente que, se estabelecemos tal lei, então ela implicará dedutivamente, junto com as condições iniciais adequadas, certos resultados ulteriores.

O modelo de cobertura por lei implica certa simetria entre explicação e previsão. A estrutura das explicações, em que deduzimos que E ocorreu a partir de condições iniciais e leis, assemelha-se à estrutura das previsões, em que deduzimos que E irá ocorrer das mesmas condições iniciais e leis. Por exemplo, se podemos explicar a chuva desta manhã pelas condições iniciais e pela lei relevante, então presumivelmente poderíamos ter previsto a chuva de antemão com base na mesma informação. Assim, pelo modelo de cobertura por lei a diferença entre a explicação e a previsão depende apenas de se você conhece o *explanandum* antes de deduzi-lo do *explanans*. Se você já conhece E, então a mesma dedução servirá para prevê-la. Uma previsão lhe diz o que esperar. Uma explicação lhe mostra que aquilo que você já conhece era de se esperar.

5.2. Explicação Teórica

Na seção anterior considerei explicações de eventos particulares, tal como a chuva desta manhã, ou certo meteoro tomar certo rumo. No entanto, o modelo de cobertura por lei é também concebido para acomodar as explicações de leis tanto quanto de eventos particulares. Por exemplo, suponha que esteja perplexo com alguma lei geral, como por exemplo, a de que sempre há arco-íris quando você olha para a chuva

com o sol num dado ângulo atrás de você. Posso explicar isso mostrando que se segue de leis que (1) a luz do sol envolve uma mistura de todos os comprimentos de onda da luz, (2) esses diferentes comprimentos de onda refratam diferentemente ao passarem da luz à água, e (3) os pingos de chuva são de forma tal que conduzirão à reflexão dentro da gota. Eis que explico uma lei com referência a outras leis. Esquemáticamente:

Explanans: L_1, L_2, \dots, L_n

Explanandum: L .

Porque o *explanandum* aqui é uma verdade geral, e não um evento particular ocorrendo num lugar e tempo específicos, não é necessário que as condições iniciais estejam envolvidas na explicação. Mas a despeito dessa diferença, ela ainda é uma explicação *dedutiva* a partir de *leis*, e por isso é ainda uma espécie de explicação “dedutivo-nomológica”. As explicações desse tipo são geralmente chamadas de “explicações teóricas” para distingui-las das “explicações particulares”.

A possibilidade das explicações teóricas mostra como o modelo de cobertura por lei pode responder uma objeção inicial comum. Considere mais uma vez a explicação particular da chuva desta manhã oferecida na seção anterior. Alguém poderia dizer que está tudo muito bem em atribuir à pressão e a umidade a chuva desta manhã, mas objetar que isso não é uma explicação até que você tenha mostrado por que as gotas em pressão em alta umidade são em geral seguidas por chuva.

O modelo de cobertura por lei pode responder insistindo que a explicação de uma chuva nesta manhã é uma coisa, e a explicação da lei de que quedas de pressão em alta umidade são seguidas por chuva é outra. Se você quer explicações de ambas, pode tê-las. Mas disso não se segue que você não tenha explicado a primeira, a chuva particular, até que tenha também explicado a segunda, a lei que dá conta da chuva particular.

Na verdade, seria obviamente autoderrotante exigir que todas as explicações contivessem explicações dos fatos aduzidos nas explicações. Cairíamos num regresso infinito. Tão logo que explicássemos as leis que originalmente apareceram no *explanans* por outras leis, teríamos, então, de explicar as outras leis por outras leis, e assim por

diante. E, no caso das explicações particulares, haveria um regresso adicional, pois precisaríamos explicar as condições iniciais mencionadas no *explanans* (por que a pressão? por que a umidade estava alta?), e isso exigiria menção a outras condições iniciais, que precisariam ser explicadas, e assim por diante.

Assim, não faz qualquer sentido exigir que numa explicação os fatos explicativos fossem sempre explicados também. Isso não é porque haja algo de errado em exigir tais explicações ulteriores em casos específicos. Acontece apenas que não podemos dar respostas a um número infinito de perguntas num tempo finito.

Que perguntas explicativas exatamente têm de ser respondidas a fim de produzir satisfação explicativa é uma pergunta interessante. Mas provavelmente não é uma pergunta que admita uma resposta geral. Se uma explicação é satisfeita depende da natureza prática daquilo que está sendo explicado a alguém. Toda gente tem uma imagem de mundo aproximadamente ampla em que eventos anteriores produzem eventos posteriores de acordo com padrões familiares. Mas alguns fenômenos não se adequam a essa imagem. O papel de uma explicação é mostrar como tais fenômenos intrigantes podem ser adequados a ela. No entanto, diferentes pessoas podem ficar perplexas com diferentes aspectos de uma situação, na qual tal situação não se adequará em suas respectivas imagens de diferentes modos. E por isso, diferentes explicações serão necessárias para satisfazê-las.

5.3. Todas as Explicações se Adequam ao Modelo de Cobertura por Lei e Vice Versa?

Nesta seção pretendo começar a levantar algumas dúvidas sobre a adequação do modelo de cobertura por lei. O modelo de cobertura por lei foi originalmente proposto por Hempel como uma análise da noção pré-analítica intuitiva de uma explicação científica. Assim, é possível perguntar se essa análise é adequada. Essa pergunta tem duas partes. (A) É verdade que toda explicação científica é uma instância do padrão de

cobertura por lei? (B) Conversamente, toda instância do padrão de cobertura por lei equivale a uma explicação científica?

5.3.1. *Todas as Explicações se Adéquam ao Modelo de Cobertura por Lei?*

Deixe-me começar com (A). Considere este exemplo. A pequena Katy pega catapora. Você quer saber por quê. Você é informado que ela brincou com Miranda, que tinha catapora. Essa parece uma explicação perfeitamente cogente. No entanto, não parece se conformar ao modelo de cobertura por lei. Suponha que pensemos que brincar com outra criança com catapora é uma condição inicial numa dedução de cobertura por lei de Katy ter pegado catapora. Então, precisamos como lei algo como “Quando uma criança que não teve catapora brinca com outra que teve, a primeira também terá”. Mas não há tal lei. Há diversos casos em que crianças não pegaram catapora após brincar como outra criança com catapora, ainda que não tivesse pegado antes.

Assim, isso é *prima facie* um contraexemplo: uma explicação intuitivamente satisfatória que não se adéqua ao modelo de cobertura por lei.

5.3.2. *Todas as Instâncias do Modelo de Cobertura por Lei são realmente Explicações?*

A pergunta (B) levantou a questão conversas: toda instância do padrão de cobertura por lei é realmente uma explicação? Eis um caso que não é.

I₁, I₂: O barômetro decaiu essa manhã; e a umidade estava alta.

L: Quando o barômetro decai em alta umidade, chove.

E: Choveu esta manhã.

Essa dedução se conforma perfeitamente às condições do modelo de cobertura por lei de uma explicação particular. Mas intuitivamente não é uma explicação satisfatória. A queda do barômetro poderia dar conta de como você sabe que vai chover. Mas estar chovendo efetivamente é um fato diferente de você saber que vai chover. E intuitivamente parece completamente errado dizer que a queda do barômetro foi responsável pela chuva.

Eis alguns casos similares.

I_1, I_2 : A sombra de um mastro M mede n ; e o sol está em um ângulo a .

L: Quando um mastro faz uma sombra de tamanho n como o sol num ângulo a , o mastro mede m .

E: O mastro M mede m .

I: A estrela E emite um desvio de luz para o vermelho.

L: Todas as estrelas como desvio de luz para o vermelho se apagam rapidamente.

E: A estrela E está se apagando rapidamente.

Ambos parecem casos impecáveis de deduções de cobertura por lei. Mas, novamente, parece completamente errado dizer que o mastro mede m porque a sua sombra mede n , ou que a estrela esta se apagando rapidamente por causa do desvio da luz para o vermelho.

5.3.3. Explicações que não são Previsões e Vice Versa

Deixe-me fazer um comentário geral sobre os dois tipos de contraexemplo ao modelo de cobertura por lei levantados nesta seção. O modelo de cobertura por lei está comprometido, como apontei anteriormente, com a idéia de que toda explicação é uma previsão potencial, e vice versa. Assim, se pudermos encontrar explicações que não são previsões potenciais, então teremos exemplos de explicações que não se adéquam ao modelo de cobertura por lei. O exemplo de Katy e a catapora são desse tipo. Você não pode prever imediatamente que ela pegou catapora só por saber que ela brincou com outra criança infectada. Porque o modelo insiste que todas as explicações deveriam ser previsões potenciais, o modelo de cobertura por lei tem problemas ao admitir essas explicações *prima facie* plausíveis.

Examinamos então o tipo converso de exemplo, previsões que de fato não são explicações e que, portanto, contaram como explicações de acordo com o modelo de cobertura por lei, mas que não deveriam contar. Você pode prever a chuva pela queda do barômetro, ou o comprimento do mastro pelo comprimento da sombra, ou o decaimento da estrela pelo desvio para o vermelho. E por isso, por aceitar que todas as previsões

potenciais são explicações, o modelo de cobertura por lei tem problemas ao excluir essas não-explicações *prima facie*.

5.4. Explicação Probabilística

Os defensores do modelo de cobertura por lei podem dar várias respostas a esses contraexemplos. Deixe-me primeiro considerar os contra-exemplos do tipo (A), designadamente, explicações intuitivamente satisfatórias, como a catapora de Katy, que não se adéquam ao modelo de cobertura por lei.

Uma possível resposta aqui seria argumentar que, se a catapora de Katy não é previsível devido a falta de uma lei que diz que ela estava segura de pegar catapora naquelas circunstâncias, então, a despeito das primeiras aparências, ela brincar com Miranda não a explica. (Afinal, outras crianças que entram em contanto com a infecção às vezes não pegam catapora. Portanto, por que supor que o contato de Katy com Miranda seja suficiente para explicar ela ter contraído a doença?) Essa via salvaria o modelo de cobertura por lei de explicação ao negar que o aparente contraexemplo fosse um exemplo genuíno de explicação.

Essa manobra, no entanto, parece nada atrativa. Seria muito estranho negar que Katy pegou catapora porque brincou com Miranda. Por isso, a maioria dos teóricos da explicação, de Hempel em diante, enfraqueceu as condições do modelo de cobertura por lei a fim de permitir que haja explicações que recorram a leis *probabilísticas* ao invés de leis sem exceções. Afinal, em nosso exemplo é presumivelmente verdadeiro que a maioria das crianças que entram em contato com a catapora a contrai, e isso significa que podemos pelo menos antecipar que Katy pegaria catapora de Miranda com uma probabilidade alta, se não com certeza. De acordo com isso, Hempel desenvolveu a seguinte modelo de “explicações *indutivo-estatísticas*” como outra espécie de explicações de cobertura por lei junto com as “explicações *dedutivo-nomológicas*”.

Condições iniciais: I_1, I_2, \dots, I_n

Leis probabilísticas: L, no sentido de que a maioria das I_1, \dots, I_n s são Es

Evento explicado: E.

As explicações que se adéquam a esse esquema são “indutivas” porque as premissas não implicam dedutivamente a conclusão, mas apenas indicam que ela tem uma alta probabilidade; e são “estatísticas” porque recorrem a leis probabilísticas ao invés de leis sem exceções.

Note que o modelo “indutivo-estatístico” de Hempel requer que o *explanans* forneça ao *explanandum* uma alta probabilidade. Não é claro que essa seja exatamente a condição correta para a explicação probabilística. Suponha que John Smith desenvolva câncer de pulmão. A explicação que damos é que ele tem fumado cinquenta cigarros por dia por quarenta anos. Intuitivamente essa parece ser uma boa explicação. Mas note que o *explanans* aqui não dá ao *explanandum* uma *alta* probabilidade. Contudo, mesmo pessoas que fumam cinquenta cigarros por dia por um longo período têm uma probabilidade baixa de desenvolver câncer de pulmão em termos absolutos. O que é verdadeiro, porém, é que elas têm uma probabilidade muito *maior* de desenvolver câncer de pulmão do que se não fumassem. Por causa disso, diversos teóricos sugeriram que a exigência de Hempel da *alta* probabilidade fosse substituída pela exigência diferente de que as condições iniciais apenas aumentem a probabilidade do *explanandum* comparada à probabilidade caso essas condições iniciais não fossem satisfeitas. (Wesley Salmon chamou a isso modelo “estatístico-relevante”, opondo-o ao modelo “indutivo-estatístico”, pois a exigência é, com efeito, que as condições iniciais fossem probabilisticamente relevantes ao *explanandum*.¹⁰)

Essa discussão da explicação probabilística levanta diversos problemas adicionais que não podem ser resolvidos aqui. Mais obviamente, você poderia querer saber se as leis probabilísticas usadas em tais explicações são supostamente reflexos do indeterminismo genuíno, ou se simplesmente refletem a nossa ignorância do conjunto completo das condições iniciais que determinam a catapora de Katy, o câncer de John Smith, etc. Diferentes respostas a esse problema conduzirão a diferentes perspectivas da explicação probabilística. Mas há pouco consenso entre os filósofos sobre como tal problema deveria ser resolvido.

10

Veja W. Salmon, *Statistical Explanation and Statistical Relevance* (Pittsburgh, 1971).

5.5. Causalidade e Explicação

Deixe-me agora considerar o outro tipo de contraexemplo, instâncias do modelo de cobertura por lei que não são de fato explicações, tal como a dedução da chuva a partir da queda do barômetro, ou da altura do mastro a partir do comprimento da sombra, ou do decaimento da estrela a partir do desvio da luz para o vermelho.

A razão óbvia pela qual essas deduções não são de fato explicações é que as condições iniciais não especificam a *causa* do evento *explanandum*. Ao invés disso, deduzem o evento *explanandum* a partir de um *sintoma* (como a queda do barômetro) ou de um *efeito* (como o comprimento da sombra, ou o desvio para o vermelho).

A solução óbvia é adicionar à abordagem de cobertura por lei a exigência adicional de que ao explicar eventos particulares, as condições iniciais devessem sempre incluir a causa do evento *explanandum*. Penso que esse é o passo correto. Mas exige diversas observações.

5.5.1. A Direção da Causalidade

Num sentido, esse passo simplesmente desvia o problema original para a análise da causalidade. Os contraexemplos do barômetro/mastro/desvio para o vermelho surgem porque as exigências originais do modelo de cobertura por lei não assegura que a “flecha do tempo”, por assim dizer, indicada nas condições iniciais ao *explanandum*. Podemos solucionar o defeito recorrendo à existência de tal seta direcionada entre eventos causalmente relacionados, e exigindo que as explicações genuínas procedam na mesma direção dessa seta.

Mas essa causalidade ter tal direção é em si uma suposição problemática. Considere a equação de Hume da causalidade com a conjunção constante. Isso por si só não nos diz, dado dois eventos constantemente conjuntados, qual está no pé da seta, isto é, a causa, e qual está na ponta, isto é, o efeito.

Assim, algo precisa ser adicionado à análise de Hume da conjunção constante a fim de adicionar a direção do tempo. Como fazer isso é uma questão bastante controversa. O próprio Hume argumentou que, dado dois eventos constantemente

conjuntados, o *anterior* é sempre a causa, e o *posterior* o efeito. Mas esse apelo à precedência temporal não é inteiramente satisfatório. (Afinal, a queda do barômetro precede a chuva, mas não a causa. E não podem algumas causas serem simultâneas com seus efeitos?)

Não proponho seguir esse problema aqui. Ainda que não seja claro como *dar conta* da direção causal, é intuitivamente claro que a causalidade *tenha* uma direção e que, exigir que as explicações sigam essa direção é a maneira de excluir os contraexemplos do barômetro/mastro/desvio para o vermelho.

5.5.2. *Todas as Explicações de Eventos Particulares são Causais?*

Não é claro que seja apropriado impor a exigência de que o *explanans* devesse mencionar uma causa sobre *todas* as explicações de eventos particulares. Suponha que explicamos por que uma substância congelada é água citando o fato de que ela é H₂O; ou suponha que explicamos por que algo tem uma temperatura *t* citando o fato de que a energia cinética média de suas moléculas é *k*. Essas são explicações razoáveis. Mas ser composto de H₂O não é a causa de algo ser água, uma vez que constitui algo ser água. Similarmente, ter energia cinética média *k* não causa a temperatura *t*, mas novamente a constitui.

Talvez essas não sejam de fato explicações no mesmo sentido em que a maioria das explicações. Elas parecem um pouco peculiares, pelo menos a mim.

Contudo, ainda que as consideremos como explicações correntes, não são de grande importância no presente contexto. Para excluir os contraexemplos do barômetro/mastro/desvio para o vermelho, precisamos exigir *alguma* conexão mais forte entre o *explanans* e o *explanandum* do que a exigida pelo modelo de cobertura por lei original. Talvez exigir uma conexão especificamente causal seja demasiado forte, pois excluiríamos assim as explicações do tipo H₂O/energia cinética média. Se assim for, então a solução é simplesmente dizer que precisamos de uma conexão *metafísica* de um tipo ou de outro em que o *explanans* ou cause ou constitua o *explanandum*.

5.5.3. *Explicações Teleológicas*

Uma razão talvez pela qual Hempel e outros proponentes iniciais do modelo de cobertura por lei fossem relutantes em impor esse tipo de conexão metafísica fosse haver uma classe importante de explicações em que o *explanans* não causa e nem constitui o *explanandum*. São as explicações *funcionais* ou *teleológicas*, explicações que desempenham um papel central na biologia, como por exemplo, em “As plantas contêm clorofila para que possam fazer fotossíntese” ou “Os ursos polares são brancos para que não possam ser vistos”. De fato, essas explicações são surpreendentes precisamente porque o item a ser explicado (a clorofila, a brancura) é a *causa*, e não o efeito, do item que explica (a fotossíntese, a camuflagem).

Se levarmos essas explicações a sério, então não está aberto a nós exigir que as explicações (não-constitutivas) sempre corram da causa ao efeito. Pois essas explicações parecem correr justamente em outra direção.

Até muito recentemente a maioria dos filósofos da ciência não levaram tais explicações a sério. Por isso, o próprio Hempel considerou as explicações teleológicas simplesmente como outro modo, junto das explicações causais normais, de exemplificar o modelo de cobertura por lei: a única diferença é que nas explicações causais o fato que explica (a baixa temperatura) precede temporalmente o fato explicado (o congelamento), ao passo que nas explicações funcionais é o fato explicado (a pele branca) que antecede temporalmente a consequência (a camuflagem) que o explica.

A maioria dos filósofos da ciência contemporâneos, no entanto, adotam um ponto de vista diferente, e argumentam que as explicações funcionais, a despeito das aparências, são realmente uma subespécie das explicações causais. Desse ponto de vista, a referência aos efeitos futuros nas explicações funcionais é apenas aparente, e tais explicações realmente se referem a causas passadas. No caso biológico essas causas passadas serão as histórias evolutivas que levam à seleção natural do traço biológico em questão. Assim, a explicação funcional da cor dos ursos polares seria entendida como nos reportando ao fato de que a sua camuflagem *passada* conduziu à seleção natural de sua brancura, e não ao fato de que eles podem ser camuflados no futuro.

Se adotarmos essa via no que diz respeito às explicações funcionais, então podemos continuar a preservar a exigência de que todas as explicações (não-

constitutivas) fluíssem da causa ao efeito, e, portanto, lidar com a dificuldade do barômetro/mastro/desvio para o vermelho do modo sugerido.

BIBLIOGRAFIA

Uma introdução clássica ao problema da indução é a seção 6 de *Os Problemas da Filosofia* (Edições 70, 2009) de Bertrand Russell. Para a resposta poperiana ao problema da indução, veja a seção 1 da *Lógica da Pesquisa Científica* (Cultrix, 1975) de Karl Popper. Uma discussão crítica das idéias de Popper pode ser encontrada nas seções 2 e 3 de *An Introduction to Philosophy of Science* (Oxford, 1989) de Anthony O’Hear. Uma abordagem fiabilista da indução é defendida na seção 4 de *Philosophical Naturalism* (Oxford, 1993). Para o “novo problema da indução” de Goodman veja a seção 3 de *Facto, Ficção and Previsão* (Editorial Presença, 1991) de Nelson Goodman e o artigo de S. Baker e P. Achinstein, “On the New Riddle of Induction”, e a resposta de Goodman “Positionality and Pictures”, em P. H. Nidditch (ed.), *The Philosophy of Science* (Oxford, 1968).

Uma boa introdução geral ao problema de distinguir leis de acidentes, tanto quanto uma defesa de sua própria posição não-humana, é dada por David Armstrong em seu *What is a Law of Nature?* (Cambridge, 1983).

A maior parte do debate contemporâneo entre realistas e instrumentalistas se foca nos argumentos em *A Imagem Científica* (Edusp, 2007) de Bas van Fraassen. Esses argumentos são discutidos depois em *Images of Science* (Chicago, 1985) organizado por P. Churchland e C. Hooker.

Há dois excelentes livros sobre o bayesianismo, os quais também fornecem uma introdução geral à teoria da confirmação e aos conceitos de probabilidade: *Probability and Evidence* (Cambridge, 1982) de Paul Horwich, e *Scientific Reasoning* (La Salle, Ill., 1989) de Colin Howson e Peter Urbach.

O ponto de partida para as discussões contemporâneas da explicação é a seção 4 de *Aspects of Scientific Explanation* (New York, 1963) de Carl Hempel. Para debates mais

recentes, veja *Explaining Explanation* (London, 1990) de David Ruben, e os ensaios em *Explanation* (Oxford, 1993) de David Ruben.